International Journal of Communication 11(2017), 1635–1660 1932–8036/20170005

# Testing the Power of Game Lessons: The Effects of Art Style and Narrative Complexity on Reducing Cognitive Bias

ROSA MIKEAL MARTEY[1]
Colorado State University, USA

ADRIENNE SHAW
Temple University, USA

JENNIFER STROMER-GALLEY
Syracuse University, USA

KATE KENSKI
University of Arizona, USA

BENJAMIN CLEGG
JAMES FOLKESTAD
Colorado State University, USA

TOBI SAULNIER
1st Playable Productions, LLC, USA

TOMEK STRZALKOWSKI
University at Albany, SUNY, USA

Rosa Mikeal Martey: rosa.martey@colostate.edu
Adrienne Shaw: adrienneshaw13@gmail.com
Jennifer Stromer-Galley: jstromer@syr.edu
Kate Kenski: kkenski@email.arizona.edu
Benjamin Clegg: Benjamin.Clegg@colostate.edu
James Folkestad: James.Folkestad@colostate.edu
Tobi Saulnier: tobi@1stplayable.com
Tomek Strzalkowski: tomek@albany.edu
Date submitted: 2015–11–14

[1] This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory contract number FA8650-11-C-7176. 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.

Educational games have generated attention for their potential to teach more successfully and with longer-lasting outcomes than traditional teaching methods. Questions remain, however, about which features of games enhance learning. This study investigates the effects of art style and narrative complexity on training outcomes of a game designed to help players mitigate three cognitive biases. The training was effective and was retained eight weeks later, although differences in art style and narrative complexity did not affect overall learning. The games were also compared with an alternative training technique, a professionally produced video. Immediately after exposure, the games produced better training than the video on two of the biases; eight weeks later, the games produced better training than the video on one of the biases.

*Keywords: educational games, serious games, art, narrative, cognitive biases*

Educational games—a type of "serious" game—have grown as an area of product development for game companies and as a subject of research to ascertain their effectiveness (Michael & Chen, 2006; Ritterfeld, Cody, & Vorderer, 2010). Of particular interest in the scholarship on educational games is their ability to teach people successfully, with learning outcomes that persist over time, especially when compared with traditional teaching methods such as training videos or lecture materials (Kirriemuir & McFarlane, 2008). Games have been widely used to train basic skills such as arithmetic, but only a few studies have looked at games' ability to teach advanced concepts such as health topics (Peng & Liu, 2008), willingness to help others (Peng, Lee & Heeter, 2010), and problem-solving skills (Li & Tsai, 2013). That research does not examine the long-term effects of playing such games, however. Another significant limitation of the literature is that, despite the existence of extensive research on the educational uses of games and some comparisons to classroom instruction (Clark, Tanner-Smith, & Killingsworth, 2014), only a few studies, discussed below, assess educational games against other visually and narratively rich media-based teaching techniques such as educational videos.

The goal of this project was to test experimentally the short- and long-term effects of narrative complexity and art style in four versions of an educational video game created for this project. All versions of the game produced effective and long-lasting training in players' ability to mitigate three different cognitive biases as measured on surveys. We compared our results to the effectiveness of an independently produced training video that teaches the same biases in order to examine how the results of game-based training may differ from other types of training. The game outperformed the comparison video immediately after the stimulus for two biases and eight weeks later for one bias.

## Effectiveness of Educational Games

Educational games have been identified as a promising tool for learning. Digital games offer new and innovative learning environments, and, as Gee (2003) and Aldrich (2005) explain, good games mirror effective learning models. As interactive media, games allow users to review educational information as well as to practice the behavior-based skills necessary for effective, long-lasting learning. Theories about games propose that they have the potential to improve learning outcomes because the interactive characteristics of the medium offer the opportunity for "learning by doing," or experiential learning (Kolb, 1984), which is thought to enhance content relevance, promote analytical reflection, and increase intrinsic motivation (Gee, 2003; McGonigal, 2011). Researchers posit that intrinsic motivation generates better learning outcomes (Deci & Ryan, 1985), and that effective games are intrinsically motivating to players (McGonigal, 2011; Salen & Zimmerman, 2003).

Overall, meta-analytic work on the impact of serious games and simulations on learning indicates that games outperform other instructional methods. For example, Sitzmann (2011) found that simulation games significantly outperformed non-game learning conditions on improvements to self-efficacy, declarative and procedural knowledge, and retention across 65 studies. Meta-analyses of 39 studies from Wouters, van Nimwegen, van Oostendorp, and van der Spek (2013) and of 69 studies from Clark and colleagues (2014) found that games (compared with nongame instructional methods) improve learning overall. The characteristics and purpose of the games themselves may be relevant to these findings. For example, Girard, Ecalle, and Magnan (2013) found in a meta-analysis of both serious and nonserious games (e.g., New Super Mario Bros) that only three of the nine studied games improved learning outcomes, and these three were serious games. The current study adds to this evidence by examining whether a game can produce meaningful, long-lasting training.

Research suggests that use of narrative in educational settings can enhance learning and engagement (Dettori & Paiva, 2009; Hinyard & Kreuter, 2007). Researchers have found that rich narratives promote reflexivity (Conle, 2003; Eisner, 1998), lead to exploratory learning (Mott, Callaway, Zettlemoyer, Lee, & Lester, 1999), provide motivating learning scenarios (Rowe, McQuiggan, Mott, & Lester, 2007), and lead to transportation and engagement (Gerrig, 1993; Green, Brock, & Kaufman, 2004). These influences are theorized to improve learning outcomes in educational settings because they provide additional contexts for comprehension that allow learners to associate instructional material with known phenomena and situated cognition, leading to improved learning transfer (Bransford, Brown, & Cocking, 2000). Only a few studies, however, have empirically tested these findings in game-based learning (Clark et al., 2014; Dickey, 2006). For example, one study compared a rich and a minimal narrative version of an educational mystery game to a PowerPoint version of the same content. They found that the rich narrative generated the lowest learning gains and that the PowerPoint version generated the highest gains (McQuiggan, Rowe, Lee, & Lester, 2008). Koenig (2007) compared an educational game with different types of narrative to a game with no narrative and found no significant differences among conditions.

Some evidence suggests that games without narratives might actually be better for learning than games with narratives (Clark et al., 2014; Wouters et al., 2013). More research is needed to determine the impact of narrative complexity on learning outcomes. The present study compares a game with named characters with backstories, significant plot events, and a broader world context to a game that minimized these features to explore whether these narrative elements contribute to learning.

The visual environment in games may also play a role in learning. In some cases, images can help learner recall and comprehension (Houts, Doak, Doak, & Loscalzo, 2006; Levie & Lentz, 1982).

Research on the impact of image complexity on learning, however, is mixed, because complex images and video can be distracting and confusing (Mayer & Moreno, 2003; Stokes, 2002). The impact of imagery in educational games is similarly unclear. One study suggests that minimalistic visuals lead players to be more engaged in games (Wolf, 2003), but others argue that rich, realistic visuals enhance feelings of co-presence (Bailenson, Yee, Merget, & Schroeder, 2006) and affect avatar credibility (Nowak, Hamilton, & Hammond, 2008), leading to greater engagement and learning. The empirical evidence is mixed. Vogel et al. (2006) found no evidence that realism (rich visuals) had an impact on learning in a meta-analysis across 32 games. In contrast, meta-analyses from Wouters et al. (2013) as well as from Clark and colleagues (2014) found that games with cartoon and with realistic visuals were less effective than the least realistic (schematic) games for learning. The latter emphasized, too, that the impact of art style and narrative are highly interrcorrelated. The present study compares a more colorful, detailed, and realistically shaded art style with a sparser, flatter, and gray-scale style to investigate these contrasting trends in the literature.

The type of learning outcome may play a role as well. For example, no improvements in learning were found in studies that compared video games to other instructional methods for teaching perspective taking and empathy (Hughes, 2014), cognitive skills (Simpson, Camfield, Pipingas, Macpherson, & Stough, 2012), or natural science and ecology (Wrzesien & Raya, 2010). One exception is a study by Peng, Lee, and Heeter (2010), who compared a game designed to generate social change, *Darfur Is Dying*, with text-based versions of its content and found that players had greater willingness to help others than those who received the text-based version.

## Cognitive Biases and Mitigation Training

The game studied in the current project aims to teach players how to reduce their susceptibility to cognitive biases. Cognitive biases arise from the human tendency to use shortcuts in information processing and decision making that can result in systematic errors or biases (Kahneman, 2011). Although research has frequently examined the triggers and manifestations of cognitive biases, few researchers have found effective ways to teach people how to overcome these deep-rooted cognitive processes (Croskerry, Singhal, & Mamede, 2014). Efforts to reduce bias in specific decision-making contexts are typically framed at the point of decision making rather than as general bias-avoidance training (Larrick, 2004), or they are based on tools that can directly influence the processing of information (e.g., decision-support systems) within the actual task environment (Arnott, 2006). Reducing individuals' overall tendencies toward bias through training was proposed by Fischhoff (1982), but implementing such training has had limited success (Sanna, Schwarz, & Small, 2002). Some bias elicitation and training techniques have been found to be reliable in nongame settings (Dale, Kehoe, & Spivey, 2007).

Games offer a promising new way of training people to be aware of and avoid cognitive biases, because the interactive and iterative characteristics of games provide opportunities to learn by doing, contextualized problem solving, and active decision making, which are promising approaches for bias mitigation (Kahneman, 2011). Furthermore, games can combine both complex and simple tasks and decisions, which allows for different types of reasoning to be applied to decision-making contexts. This situated learning may be particularly effective for debiasing training (Stewart, Latu, Kawakami, & Myers, 2010). The use of these debiasing techniques in a game environment has not previously been evaluated.

This project focused on three cognitive biases identified in decision-making and psychological literature as among the most significant (Kahneman, 2011). Fundamental attribution error (FAE) is the tendency to assume that dispositional or personality-based factors rather than situational factors account for the behavior of others (see Gilbert & Malone, 1995). Confirmation bias (CB) is the tendency to seek out and remember information that matches or supports one's view (see Nickerson, 1998). As a type of selective exposure, this bias often results in poor-quality or uninformed decisions that reinforce previously held beliefs (Fischer & Greitemeyer, 2010). Bias blind spot (BBS) is the tendency to consider one's own decisions as being free from cognitive bias, even when one can recognize that bias in others (see Pronin, 2007). This notion informs communication research on, for example, the third-person effect (Sun, Pan, & Shen, 2008).

Because research on the impact of specific narrative and art styles in educational games is inconclusive and often in conflict, we use research questions rather than hypotheses in this study. We asked:

*RQ1: Can an educational digital game result in significant bias reduction of three cognitive biases: fundamental attribution error, confirmation bias, and bias blind spot?*

*RQ2: Does delivery of cognitive bias training via a digital game result in better bias reduction than does comparable content delivered via an educational video?*

*RQ3: Does a rich narrative produce better bias reduction between pre- and posttests than does a light narrative?*

*RQ4: Does detailed art style produce better bias reduction between pre- and posttests than does minimal art style?*

In addition, for an educational game to have a lasting impact, players must retain training over time. Thus, this project investigated two research questions on the long-term effects of the training:

*RQ5: Can an educational digital game result in significant reduction of the three biases eight weeks later?*

*RQ6: What are the effects of art style and narrative on reduction of biases eight weeks later?*

## Methods

To examine the impact of art style and narrative on bias mitigation, we created four versions of an educational game that varied art style and narrative complexity. We used a 2 × 2 full factorial experiment—rich narrative (RN) versus light narrative (LN) and detailed art style (DA) versus minimal art style (MA). To compare the games' training outcomes to those from material likely to be used in a training setting, we also tested the impact of an independently produced video that trains the same content but with different format and language than used in the game.

Two critical elements were incorporated into the study design to demonstrate the effectiveness of training. First, learning and bias avoidance were evaluated outside of the game environment via measures described below. This was done to assess transfer from training to other settings (Holding, 1991), which is a key requirement of an effective learning process. Second, some forms of training and instruction maximize performance immediately after exposure, whereas learning is best considered from the longer-term retention of knowledge (see Schmidt & Bjork, 1992); the psychological literature extensively documents trainees inability to retain newly learned knowledge and skills (e.g., Annett, 1979). Therefore, mitigation was examined both immediately after and at eight weeks after the training with no additional refresher training.

### *Stimuli*

The game is a Flash-based puzzle game in which players navigate an avatar through a series of nine rooms that teach about the three biases and how to mitigate them. Each room has an introduction that explains the bias, a series of puzzles to be solved, and a summary of the room's lesson in a separate transition area. A help system provides additional instructions and hints for players who are stuck. To minimize the potential effect of disidentification in the game, the player avatar is as gender and racially neutral as possible (Shaw, 2014) with neutral clothing and a helmet to obscure face and hair.

Art style was manipulated by keeping the perspective and functionality of the game the same, but varying color, shading, and objects. The detailed art style version is full-color with rich texture and shading and realistic detailing. The minimal art style condition is monochrome with minimal shading and almost no textures (see Figure 1). There are no interactivity differences between the two games. The detailed art condition was designed to include the "seductive details" and more realistic object rendering theorized to have an impact on learning by enhancing engagement; the MA condition aimed to minimize these because some research suggests that extraneous details reduce learning (Mayer & Moreno, 2003).

For each of the art style conditions, two separate narrative versions of the game were built. In the light narrative condition, players are told only that they were in a training center to learn about cognitive biases. The rich narrative version positioned players as the child of a bias reduction expert who once worked with the CYCLE Center owner, Dr. Ohm. In this narrative, the player infiltrates the center with the help of its top trainer, Tallie, in order to learn Ohm's techniques and share them with the world. In the process, the player discovers that Dr. Ohm is using human brains to create robots and vows to shut down the evil operation. These narrative details provide a sense of backstory, robust characters, and a world in which significant plot events occur in order to leverage the narrative engagement and transportation theorized to enhance learning (Green et al., 2004).


***Figure 1. A room in the CYCLES game with minimal art (above) and detailed art (below).***

Much of the rich narrative story is told in an introduction, in text added to transition rooms, and in two possible game endings that depended on choices made in the final room. The puzzle room text is almost dentical across the narrative conditions to maintain equivalent content delivery. The voice-over was adjusted for the delivery of specific plot lines in the rich narrative condition to be more dramatic and character-driven. Aside from the introduction and conclusions of the games, which used version-specific informational screens, each art condition is the same across narrative conditions.

The comparison video is a high-quality film created independently by the video production company 522 Productions. It was professionally written, acted, and produced with consulting by outside experts in cognitive biases. It uses the same definitions and types of examples (although not exactly the same examples) for the biases as the game. The video presents a professor character who presents and discusses a series of scenarios in which people exhibit cognitive biases. Through humor and real-world examples, the video aimed to maintain interest and engagement. This condition was not conceptualized as a traditional control, but rather as a comparison to training the same material using a different medium to determine whether an alternative training technique likely to be used in classrooms has different learning outcomes. A video was used because it is visually and narratively complex and engaging, but does not allow for learner interaction.

### *Study Procedures*

Participants were recruited from college classes and psychology subject pools at three universities in the United States in fall, 2012. Recruitment materials made no mention of video games. Participants were scheduled to come to a campus computer lab for 2–3 hours, where they answered a questionnaire on the computer using the survey software Qualtrics. They were randomly assigned to play a version of the game (36 minutes mean play) or watch the video (30 minutes). Immediately afterward, they took the posttest survey. They were compensated with class credit. Eight weeks after they played, particilants received a third survey via an email link. They were paid a $10 gift card for completing this final survey. Surveys assessed computer and game experience, measures of the three biases, demographics, and psychological variables including Need for Cognition and the Big Five personality assessment.

### *Study Participants*

There were 480 subjects in the final data set after excluding 16 participants for errors in data collection, and 184 of these subjects (40%) completed the retention test. The sample was 65% female and 35% male, and 79% were between ages 18 and 20. Slightly over half (54%) of participants reported they have intermediate-level computer skills, 23% advanced, 10% expert, and 13% novice. Most participants did not consider themselves “gamers” (84%), and 20% reported playing games less than once every few weeks, 19% every few weeks, 17% for 1–2 days a week, 14% for 3–5 days a week, 30% at least once a day.

### *Measures*

We developed three new scales to measure the cognitive biases that were the focus of study (FAE, CB, and BBS; for details, see Clegg et al., 2014). All three scales were transformed to a −100 to +100 scale, with 0 indicating no bias, allowing us to assess whether mitigation resulted in tendencies toward the opposite bias (e.g., disconfirmation bias).

*Fundamental attribution error scale.* This measure presented participants with 10 brief scenarios, such as

> One of your peers receives an A in a course that has a reputation for being hard. The best explanation for this student's grade is that the student is smart. To what extent do you agree or disagree with this assessment?

Scenarios were based on a range of topics areas, such as donating money, romantic relationships, and driving. Answers used sliders on a scale of 1 (*disagree*) to 7 (*agree*) and were scored by taking the mean of all 10 scenarios. The FAE scale was reliable with Cronbach alphas of .70 (pretest), .89 (posttest), and .88 (retention).

*Confirmation bias scale.* To examine CB, we tested how participants weight information in relation to a known hypothesis, based on the paradigm developed by Cook and Smallman (2008). For this measure**,** each participant is presented with seven scenarios, such as "You are considering taking a trip to the country Calzycoah. You don't speak the language, but you will not have to pay for airfare to this vacation spot." Four of these scenarios then provided a hypothesis, three asked participants to decide between two possible hypotheses. Participants then rated six pieces of evidence about each scenario for the importance of "asking the following questions in evaluating your selected decision"; three represented confirming evidence, and three disconfirming evidence. Answers used sliders on a scale of 1 (*unimportant*) to 7 (*extremely important*). Scores were calculated by subtracting the mean of the confirming items from the mean of the disconfirming items. The CB had reliabilities of alphas of .82 (pretest), .77 (posttest), and .75 (retention).

*Bias blind spot scale.* BBS was measured with a variant of the method employed by Pronin, Lin, and Ross (2002). Participants were first asked to rate themselves and an "average student" at their institution on seven different positive characteristics and traits. A second set of items then informed them of the *illusion of superiority*, the tendency to rate oneself as above average on these types of dimensions. They were then asked, "To what extent do you believe that you showed this tendency when you rated your [intelligence] on a previous question?" This was followed with: "To what extent do you believe that the average student from your university would show this tendency if he or she rated his or her [intelligence]?" Answers were on sliders from 1 (*not at all*) to 9 (*very much*) and were scored by subtracting ratings of others from ratings of the self. The scale had Cronbach alphas of .88 (pretest), .85 (posttest), and .85 (retention).

Although the FAE pretest reliability is slightly low, it is not unusual to show pretest low reliability for topics about which learners have very little knowledge (Bacon, 2004) because it is not expected that they would have consistent responses to such items (Reichardt, 1979). More generally, pretest reliabilities of .70 are noted as common and acceptable for basic research (Cohen, 1977), especially with performance assessments (Bacon, 2004; Sax, 1997).

The bias measures were developed using extensive pretesting. Twenty participants who were not part of the main experiment were given a pre- and posttest on the bias measures with no feedback between administrations. The two attempts at the questions were separated by an intervening distractor task (mathematical problems). Correlations between pre- and posttest suggest sufficient reliability for all measures of bias across the tests (test 1 to test 2 correlations, FAE: $r = .94$; CB: r = .69; BBS: $r = .72$). The test–retest data also were used to examine whether practice effects occurred for the bias measures. None were found, FAE: mean (*SD*) test 1 = 10.4 (24.2), test 2 = 12.5 (25.6), $t(19) = 1.10$, $p > .25$; CB: test 1 = 24.4 (13.8), test 2 = 24.7 (11.1), $t(19) = 0.07$, $p > .90$; BBS: test 1 = 13.2, test 2 = 12.4 (12.3), $t(19) = 0.41$, $p > .65$). All scores remained stable across the two administrations of the instruments, with no evidence of significant decreases in bias from exposure to the bias questions alone.

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

## Results

### *Analytical Approach*

To examine whether a video game can train users to mitigate cognitive biases (RQ1), we used a repeated measures multivariate analysis of variance comparing the three bias measures at pretest with immediate posttest for the game condition only ($n = 391$). To examine RQ2 through RQ4, we examined all conditions ($N = 480$) using a pair of analyses: first, one featuring the repeated measures bias score from pretest and posttest by the five conditions. Second, we examine posttest-only performance by condition using planned linear contrasts to explore the training outcome effects of game play versus the video, differences between the art style conditions, and differences between the narrative complexity conditions. The same approach is used to examine RQ5 and RQ6 comparing pretest to the eight-week posttest ($n = 184$ with video; $n = 144$ games only). Because BBS analyses revealed some complex effects, we also conducted planned comparisons to assess the pre–post and pre–retention effects on this bias of each condition. Table 1 provides the means and standard deviations in each condition for each bias.

***Table 1. Immediate and Delayed Bias Mitigation Descriptive Statistics.***

| | FAE pretest | FAE posttest | FAE retention | CB pretest | CB posttest | CB retention | BBS pretest | BBS posttest | BBS retention |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Light Narrative/Minimal Art | | | | |
| *n* | 99 | 99 | 30 | 99 | 99 | 30 | 99 | 99 | 30 |
| *M* | 21.2 | −15.9 | 4.3 | 24.3 | 13.2 | 12.2 | 14.1 | 10 | 8.7 |
| *SD* | 27.7 | 42.3 | 38.9 | 16.9 | 20.4 | 17.3 | 20.4 | 19.9 | 21.1 |
| | | | | | Light Narrative/Detailed Art | | | | |
| *n* | 93 | 93 | 36 | 93 | 93 | 36 | 93 | 93 | 36 |
| *M* | 25.9 | −12.7 | 5.1 | 26.4 | 14 | 12.4 | 14.7 | 14 | 13.2 |
| *SD* | 25.2 | 37.8 | 25.6 | 19.1 | 20.1 | 17.4 | 19.9 | 21.5 | 17.0 |
| | | | | | Rich Narrative/Minimal Art | | | | |
| *n* | 95 | 95 | 40 | 95 | 95 | 40 | 95 | 95 | 40 |
| *M* | 23.2 | −20.5 | 3.2 | 21.9 | 10.2 | 12.7 | 14.1 | 8.4 | 10.8 |
| *SD* | 27.3 | 41.3 | 40.3 | 16 | 17.1 | 18.9 | 19.6 | 15.6 | 19.2 |
| | | | | | Rich Narrative/Detailed Art | | | | |
| *n* | 105 | 105 | 38 | 105 | 105 | 38 | 105 | 105 | 38 |
| *M* | 26.4 | −13 | −3.3 | 22 | 10.4 | 10.8 | 14.2 | 13 | 8.0 |
| *SD* | 23.7 | 44.4 | 39.3 | 19.6 | 21.8 | 21.1 | 22.4 | 22.7 | 17.1 |
| | | | | | Video | | | | |
| *n* | 89 | 89 | 40 | 89 | 89 | 40 | 89 | 89 | 40 |
| *M* | 24 | 9.9 | 17.6 | 21.7 | 22.8 | 16.7 | 10.3 | 11.8 | 5.2 |
| *SD* | 25.5 | 36.8 | 28.1 | 19.1 | 20.9 | 19.4 | 21.5 | 18.8 | 22.8 |

*Note.* FAE = fundamental attribution error; CB = confirmation bias; BBS = bias blind spot.

### ***Bias Reduction at Immediate Posttest***

Overall, across all biases, we found a significant effect of training from pre- to immediate posttest (RQ1) for the game conditions only, Wilks's $\lambda$ = .46, $F(3, 385) = 151.71$, $p < .0005$, indicating that people learned from the games. There was no difference in the interaction between condition and pre–post scores examining the training effects across the four game conditions, Wilks's $\lambda$ = .99, $F(9, 937) < 1$. This omnibus analysis provides strong evidence that games improve learning (RQ1), showing overall effectiveness of the games as a training platform with a large effect size (1 − Wilks's $\lambda$ = 54% of variance accounted for). In Figures 2 through 7, bars represent mean values by condition, and error bars show standard error.

*Fundamental attribution error training.* Examining FAE for individual conditions, we found that training improves learning (RQ1; see Figure 2), with effective training seen in a main effect of pre–post, *F*(1, 475) = 393.18, $p < .0005$, *MSE* = 726, and a main effect of condition, *F*(4, 475) = 4.28, $p < .05$, *MSE* = 1,604). There was an interaction between pre–post and condition, *F*(4, 475) = 8.47, $p < .0005$, demonstrating that the game was better than the video. Posttest performance was examined using planned linear contrasts. These showed that the game trained FAE better than the video (RQ2: all game conditions vs. video), *t*(475) = −5.26, $p < .0005$). There was no difference between either the game narrative conditions (RQ3), *t*(475) = 0.48, $p > .05$, or the game art style conditions (RQ4; *t*(475) = −1.41, $p > .05$).

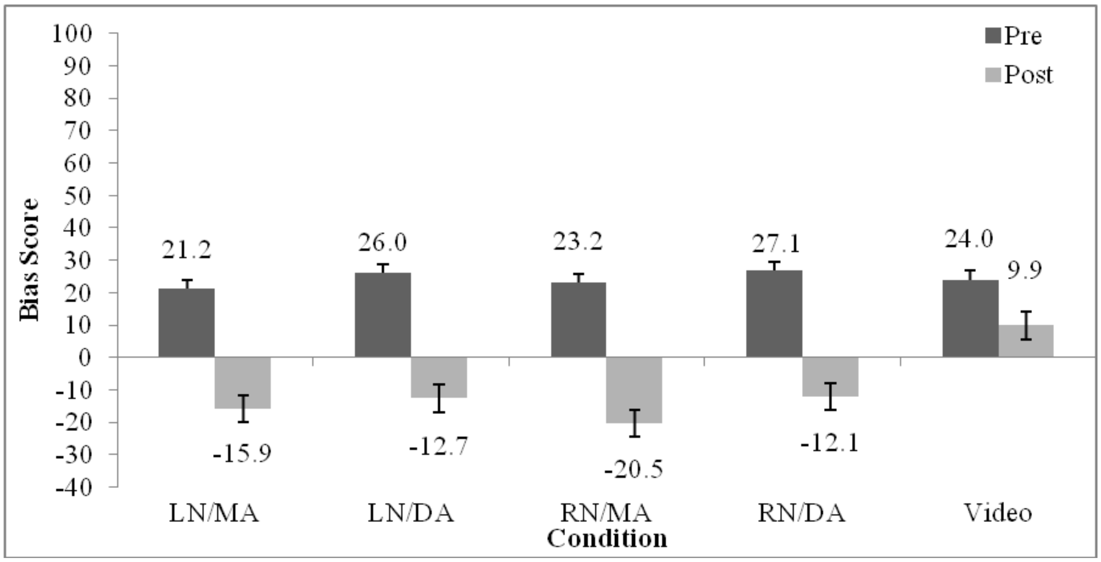


***Figure 2. Fundamental attribution error by condition for pretest and posttest.***

*Confirmation bias training.* CB was improved with training (RQ1; shown in Figure 3), demonstrated by a main effect of pretest–posttest, *F*(1, 475) = 107.41, $p < .0005$, *MSE* = 187, a main effect of condition, *F*(4, 475) = 2.47, $p < .05$, *MSE* = 550), and differential effects of training seen in an interaction between pre–post and condition, *F*(4, 475) = 8.11, $p < .0005$. This demonstrates that the game was better than the video. Planned linear contrasts showed an advantage for game play versus the video (RQ2: all game conditions vs. video), *t*(475) = −4.61, $p < .0005$. Game narrative was marginally nonsignificant (RQ3; *t*(475) = 1.66, $p = .09$), with a trend toward better training from the minimal narrative games. No difference was observed between the game art style conditions (RQ4; *t*(475) = −0.19, $p > .05$).

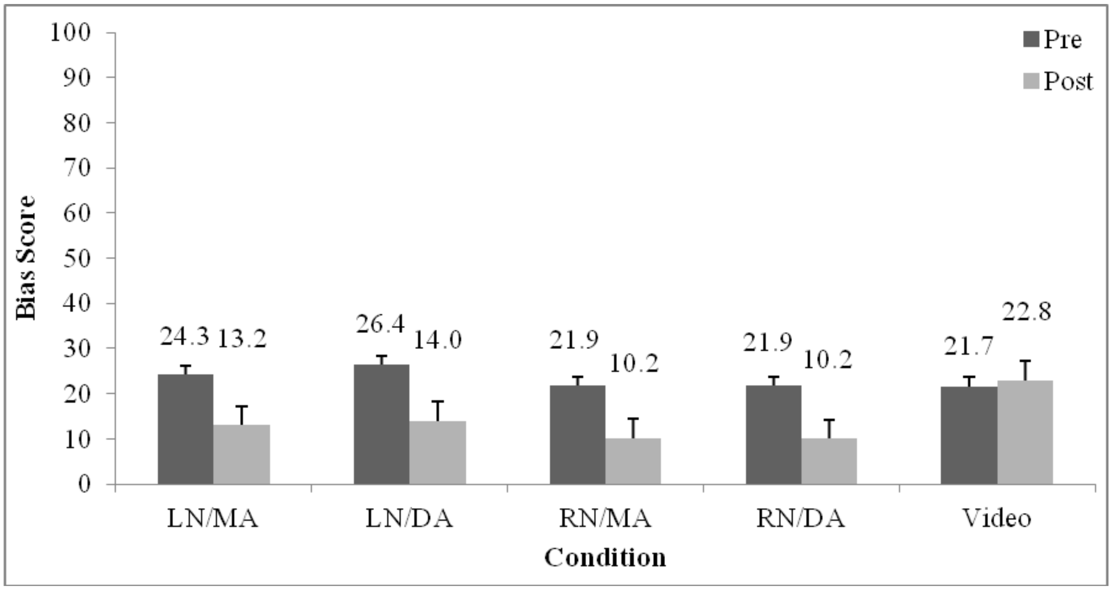


***Figure 3. Confirmation bias by condition for pretest and posttest.***

*Bias blind spot training.* BBS showed no strong overall change from training (RQ1; see Figure 4), demonstrated by a marginally nonsignificant main effect of pre–post, $F(1, 475) = 3.71$, $p = .055$, $MSE = 268$, no main effect of condition, $F(4, 475) < 1$, $MSE = 398$, and no interaction between pre–post and condition, $F(4, 475) = 1.40$, $p > .05$. However, analysis showed a significant improvement in BBS overall from the aggregated game conditions when excluding the video control condition, $t(389) = 2.41$, $p < .05$, $d = .12$, although the size of the effect was relatively small.

Post hoc planned linear contrasts showed no overall advantage in training BBS for game play versus the video (RQ2: all game conditions vs. video), $t(475) = -0.19$, $p > .05$. There was a difference between the art style conditions with greater mitigation of bias associated with games featuring more detailed art (RQ4; $t(475) = -2.10$, $p < .05$). There was no difference between the narrative conditions (RQ3; $t(475) = 0.67$, $p > .05$). Planned comparisons revealed that only one game condition—rich narrative/minimal art—showed significant reduction in bias from pre- to posttest, $t(94) = 2.67$, $p < .01$.

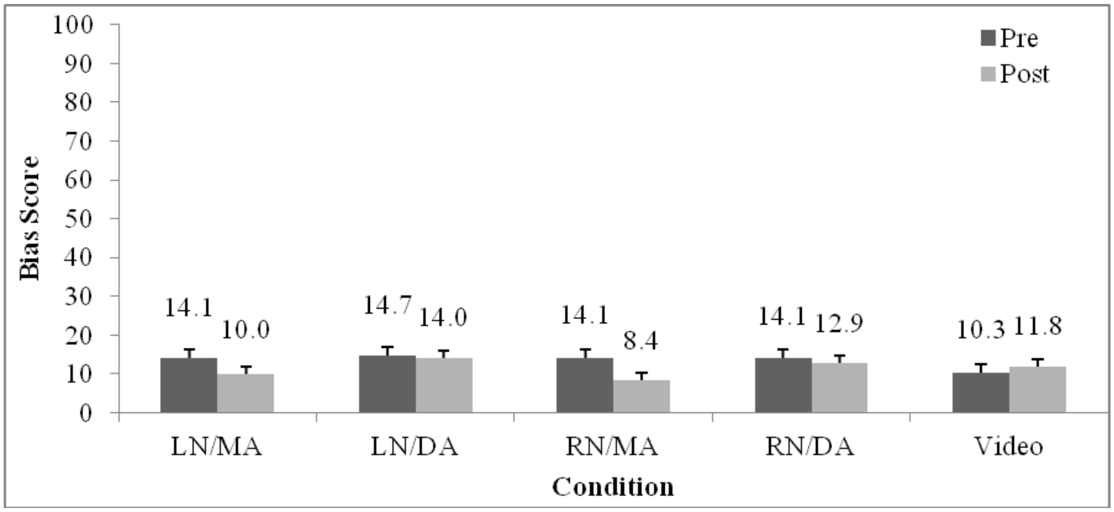


***Figure 4. Bias blind spot by condition for pretest and posttest. Results at eight-week retention test.***

A repeated measures analysis of variance comparing pretest bias to the eight-week retention bias across the three types of bias as dependent variables (RQ5) revealed that, overall, significant training was retained by those who played the games ($n$ = 144), Wilks's $\lambda$ = .62, $F(3, 138) = 28.04$ $p < .0005$, but there was no difference among game conditions, $F(9, 336) < 1$. This omnibus analysis provides evidence for the overall effectiveness (RQ5) of the games for retaining prior training (with a medium effect size, showing 38% of variance accounted for).

*Fundamental attribution error training.* Examining FAE across all conditions showed improvements retained eight weeks after training (RQ6; see Figure 5), with a main effect of pretest to retention bias levels, $F(1, 179) = 51.06$, $p < .0005$, $MSE = 557$, suggesting that training was being retained. There was no main effect of condition, $F(4, 179) < 1$, $MSE = 1{,}350$, but there was an interaction between pre–retention and condition, $F(4, 179) = 3.05$, $p < .05$, showing differences between the conditions in the amount of training retained. Examining the retention test performance using the planned linear contrasts showed an advantage for game play versus the video (all game conditions vs. video), $t(179) = -2.45$, $p < .05$, no difference between the game art style conditions, $t(179) = 0.49$, $p > .05$, and no difference between the game narrative conditions, $t(179) = 0.82$, $p > .05$.

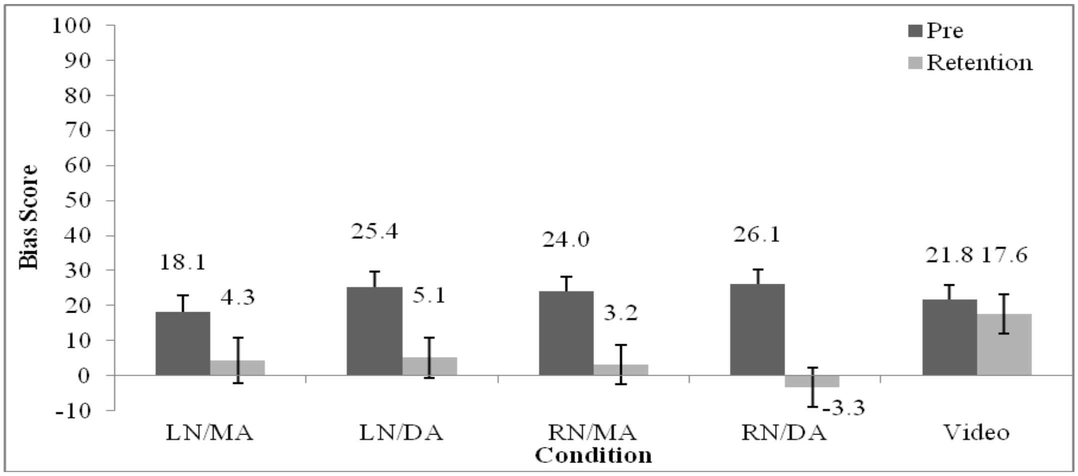


***Figure 5. Fundamental attribution error by condition for pretest and retention test.***

*Confirmation bias training.* Performance on CB was retained (RQ6; see Figure 6), demonstrated by a significant main effect of pretest to retention test, $F(1, 179) = 50.59$, $p < .0005$, $MSE = 227$. There was no main effect of condition, $F(4, 179) < 1$, $MSE = 462$, and no interaction between pre–retention and condition, $F(4, 179) = 1.25$, $p > .05$. Planned linear contrasts showed no significant advantage for all game play versus the video (all game conditions vs. video), $t(179) = -1.37$, $p > .05$, no difference between the game art style conditions, $t(179) = 0.25$, $p > .05$, and no difference between the game narrative conditions, $t(179) = 0.17$, $p > .05$.

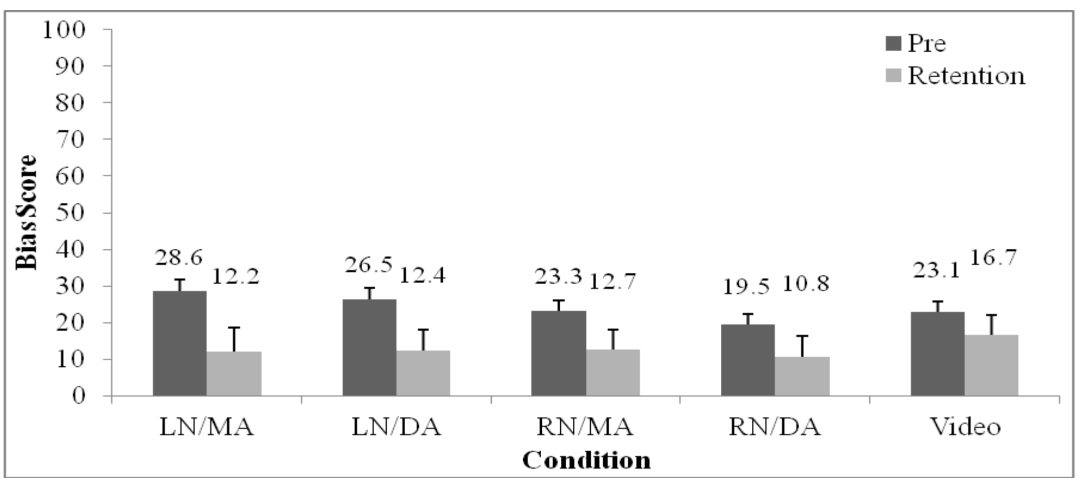


***Figure 6. Confirmation bias by condition for pretest and retention test.***

*Bias blind spot training.* BBS found that training was retained eight weeks later with a main effect of pre–retention, $F(1, 179) = 6.09$, $p < .05$, $MSE = 374$, no main effect of condition, $F(4, 179) < 1$, $MSE = 380$, and no interaction between pre–retention and condition (RQ6; see Figure 7), $F(4, 179) < 1$. Post hoc planned linear contrasts showed no significant advantage in BBS for the games versus the video (all game conditions vs. control video), $t(179) = 1.43$, $p > .05$, no difference between the game art style conditions, $t(179) = -0.27$, $p > .05$, and no difference between the game narrative conditions, $t(179) = 0.46$, $p > .05$.

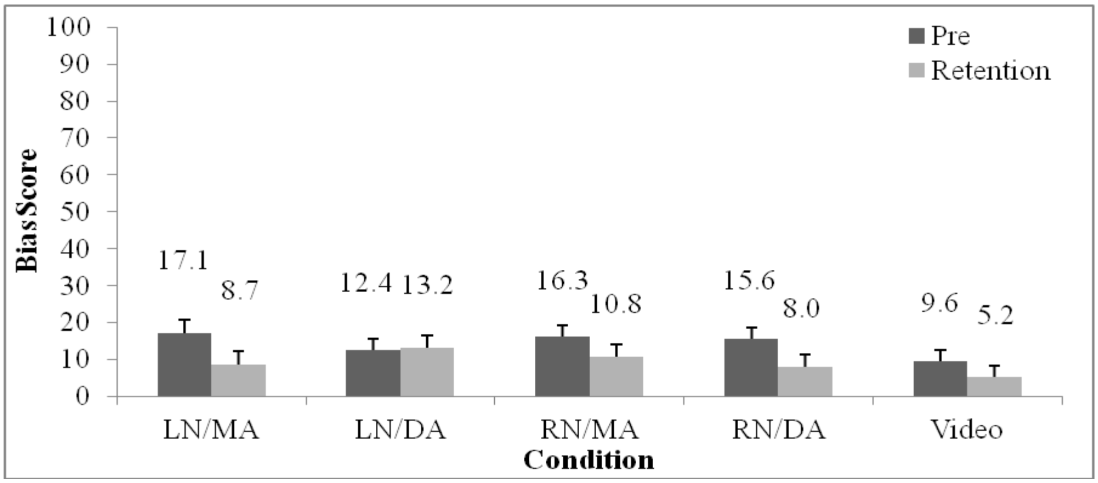


***Figure 7. Bias blind spot by condition for pretest and retention test.***

*Effect sizes.* To interpret the size of the learning effects, we also provide effect sizes and percentage change for pre- to posttest and pre- to retention-test bias levels. Effect sizes were determined using Cohen's *d*, which is calculated as the mean of the pretest minus the post- (or retention) test divided by the standard deviation of this difference score, shown in Table 2.

***Table 2. Effect Size (d) by Condition for Pretest to Posttest and Pretest to Retention-Test Effects.***

| Condition | FAE pretest/ posttest | CB pretest/ posttest | BBS pretest/ posttest | FAE pretest/ retention | CB pretest/ retention | BBS pretest/ retention |
|---|---|---|---|---|---|---|
| LN/MA | 0.91 | 0.56 | 0.18 | 0.36 | 0.88 | 0.33 |
| LN/DA | 1.06 | 0.66 | 0.03 | 0.71 | 0.56 | −0.03 |
| RN/MA | 1.06 | 0.66 | 0.27 | 0.57 | 0.45 | 0.23 |
| RN/DA | 0.96 | 0.47 | 0.05 | 0.85 | 0.44 | 0.26 |
| Video | 0.49 | −0.09 | −0.08 | 0.14 | 0.35 | 0.14 |

*Note.* FAE = fundamental attribution error; CB = confirmation bias; BBS = bias blind spot; LN = light narrative; RN = rich narrative; MA = minimal art style; DA = detailed art style.

To determine percentage reduction in bias, individuals' mean bias scores were constrained to be zero or greater because raw scale scores can include negative values. A bias percentage score was calculated as 100 − bias mean / 100. Percentage reduction (see Figures 8 and 9) was calculated as: (posttest [or retention test] % score − pretest % score) / (1 − pretest % score).

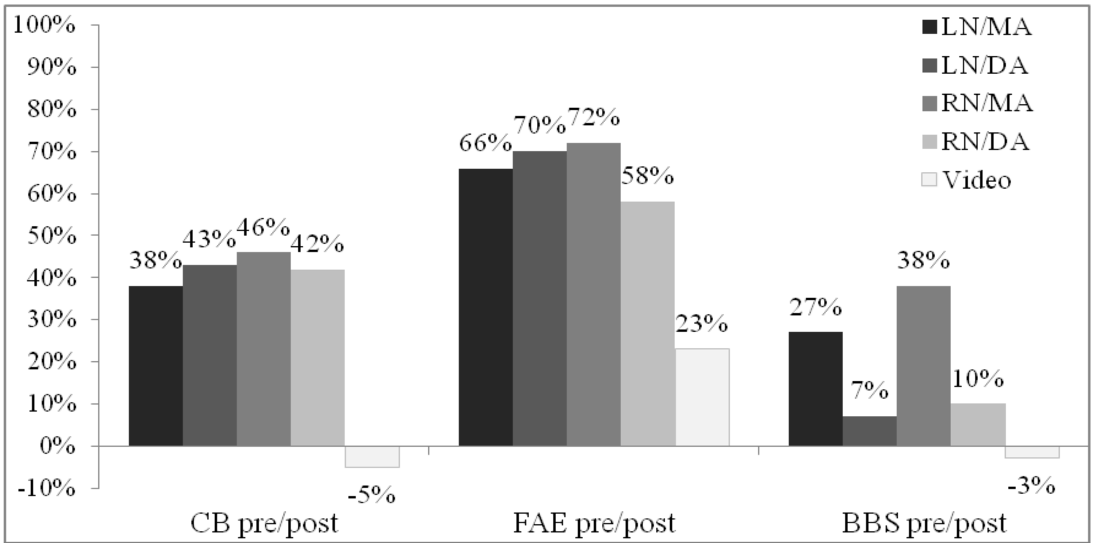


***Figure 8. Percentage reduction in each bias from pre- to posttest by condition.***

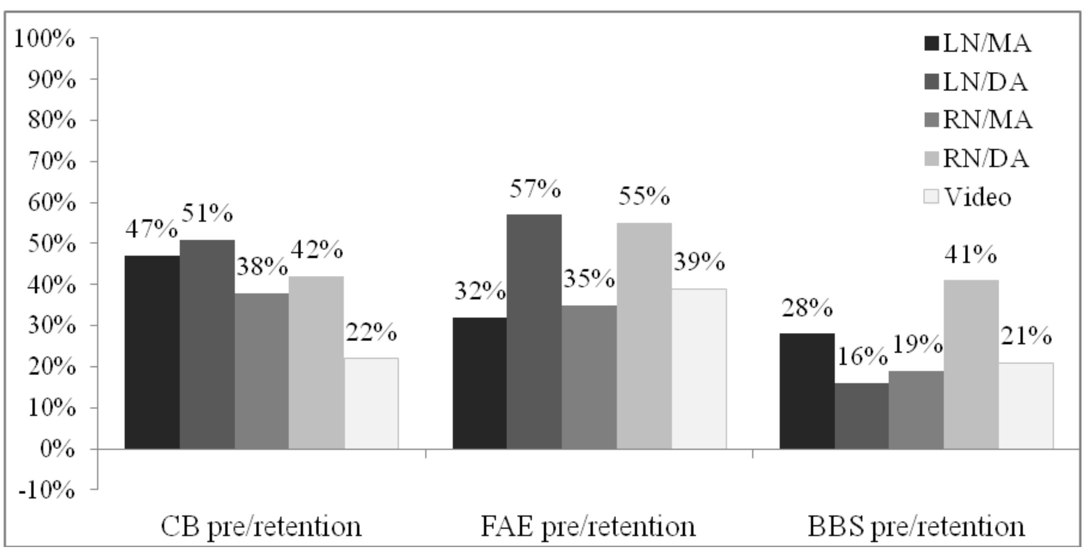


***Figure 9. Percentage reduction in each bias from pretest to retention test by condition.***

## Discussion

The results of the experiment suggest that an educational game that teaches people how to mitigate cognitive biases performed better than a training video immediately after game play (RQ1, RQ2), and that training effects lasted eight weeks later (RQ5), although the game's advantage over the video dropped away over time. We did not find that art style or narrative complexity had a significant impact on training immediately after training (RQ3, RQ4) or eight weeks later (RQ6). Immediate posttest performance in bias blind spot mitigation for the minimal art style conditions was better than rich art, but effects were very small. The literature also does not provide clarity about the role of narrative or art style on learning. Scholars have tended to assume that both of these factors matter in game spaces, but our experimental results do not suggest that these have an impact on learning in games.

It could be that the way we embedded narrative—primarily at the beginning and end of the game—may have effectively created little difference for the participants as they experienced the game play itself. These findings parallel those of Sitzmann (2011) and Clark et al. (2014), whose meta-analyses found that learning outcomes are not significantly influenced by the entertainment value of games. Analyses of bias blind spot and percentage changes in bias levels hinted that art style might have an impact on training, but these may have been too small to be revealed with our sample size. Additional tests of the impact of these factors are needed. It may be, too, that the short length of these games (averaging 36 minutes) may not be sufficient to leverage the potential impact of narrative or art styles on learning. The deep engagement with a storyline generally considered the most effective in learning contexts may not have been feasible in such a short time. Indeed, although the greater detail such as color, additional objects, characters, and backstory in the game's rich art and narrative style conditions may have initially been distracting. Over time, these factors may generate the beneficial reflexivity, exploratory learning, and motivation associated with narrative in learning contexts (Dettori & Paiva, 2009; Hinyard & Kreuter, 2007).

The effect sizes of pre- to posttest and from pre- to retention test for confirmation bias and fundamental attribution error are moderate to large, suggesting that training had a substantial impact on participants' ability to mitigate these biases. Even eight weeks later, some games generated an improvement in mitigation of fundamental attribution error and confirmation bias of over 50%. Although some of the differences between the games and the video could be due to minor content differences, differences in effect sizes are large enough to suggest that a game format offers more effective learning than a video. It is likely that these differences are influenced by the game's ability to provide frequent testing in the form of in-game quizzes; such testing is known to greatly improve recall and comprehension (Rowland, 2014). Thus, we conclude that the affordances of digital games as a medium can generate greater learning outcomes, but more research is needed to assess the contribution of individual affordances in isolation. Future research should further examine how specific elements of games—such as quizzes, interactivity, and learner control over pacing—affect learning outcomes when compared with content without those elements such as videos.

The greatest improvement on mitigation was in fundamental attribution error, with more than 55% improvement in two conditions, and large effect sizes of close to 1 on Cohen's *d* for all game conditions at immediate posttest. At the retention test, we saw the expected decay in training over time, but mitigation improvement remained above 50% for the two minimal art style conditions. This may be due to the fact that, compared with CB, FAE in the games used simpler and more concrete concepts for training mitigation strategies. The effective FAE training from the video, although smaller than the game, suggests also that FAE may be easier to train against than the other biases.

The games improved confirmation bias mitigation between 38% and 46% with moderate effect sizes at immediate posttest, and between 38% and 51% at retention (see Table 2). Although effects were not as large as for the FAE training, these results represent important and substantial improvements in mitigating this pernicious bias, especially at retention. The complexity of CB and its application make communicating mitigation strategies more difficult, and the puzzles used for CB were necessarily more nuanced than others in the game. The most substantial challenge seems to be helping players understand how to translate CB mitigation to new contexts. To address this, we used several different settings for CB training, including real-world scenarios presented via in-game quiz questions.

In the two light narrative conditions, average CB mitigation was actually *better* at the eight-week retention test than it was at the immediate posttest. This may be because participants were able to apply CB mitigation to other contexts in their lives during those months, thus broadening their understanding of this concept. This suggests that training might be more effective when a person *experiences* CB and then understands how the resulting decisions are subject to bias. The game was designed to encourage players to actively commit CB and then learn to avoid it, which may also explain why the game did better than the video at training against this bias. This is supported by the finding that video participants were better on average at mitigating *all* biases at the eight-week retention test than at the immediate posttest, although the FAE effect size was smaller at retention. Because video participants did not have the opportunity to experience the biases and act on these lessons until after the session, their full comprehension may have been delayed until the retention test. This pattern suggests that using a game for training may help players actively apply lessons as they learn, but more passive methods such as videos require time for learners to integrate the ideas presented.

Although combined, the games showed significant effects of bias blind spot training, planned comparisons showed that it was only reduced in one of the five conditions by 38% at immediate posttest (likely due to the lower variance in that condition). BBS was not reduced in any conditions at retention. This may be because mitigation of this bias is more challenging than others. As Pronin (2007) and others have found, BBS can actually *increase* among those with greater knowledge of biases or greater cognitive sophistication (West, Meserve, & Stanovich, 2012). This translates to significant challenges in training against BBS while simultaneously addressing other biases. Pronin's work suggests that this effect is in part due to an increased awareness of biases leading to increased confidence that biases can be completely avoided.

Overall, these results suggest that the strength of using games to address bias mitigation is related to the complexity of the concepts required to understanding the phenomena and the time people need to consider the elements of biased behavior. Learning about biases that are easier to grasp, such as FAE, may not require the interactive, exploratory contexts that games offer—although such contexts do improve learning. Biases people find more difficult to understand and see in their own behavior, in contrast, may be better addressed with games that provide time and engaged activities. This phenomenon may also be related to the training content: Complex concepts such as cognitive biases may simply take more active exploration and thought to learn than knowledge-based topics.

This study had some limitations. Because there did not previously exist widely used scales to measure these biases in surveys, we had to develop our own. Although test–retest reliability and validation techniques did not show evidence of problems, additional research is needed to further validate these measures. Another limitation is the use of a training video for comparison rather than a traditional control. Although the video was designed to address the same topics as the game, it necessarily used different language, visuals, narrative, and topics than the games. It was also a few minutes shorter than the average game-play time. Future research should further compare the impact of training games against alternative teaching techniques. In addition, because our sample was nonrandom and drawn from university undergraduate students, our results are not generalizable to broader adult populations. It is possible that being a student regularly engaged in a variety of learning contexts changed responses to the training. Further research with different populations is needed.

## Conclusion

The burgeoning research on educational games suggests that such games might impart learning on a variety of subjects more effectively than traditional teaching methods (Aldrich, 2005; Gee, 2003). In this experiment, a video game outperformed a training video with similar content in helping participants learn to mitigate fundamental attribution error and confirmation bias, and this training was retained eight weeks later, although the advantage of the game for confirmation bias training dropped away. Video participants were actually *better* at mitigating the biases eight weeks later than they were immediately after training. Results from the experiment also suggest that art style and narrative complexity differences did not influence training effectiveness.

Overall, our experiment suggests that educational games do hold promise for providing persistent training in avoiding cognitive biases, a fairly pervasive aspect of human information processing and decision making. Moreover, this study shows that this training can have moderate to large effects on mitigation for at least two of the biases we examined. The combination of information delivery and practice led to lasting and significant training effects. The fact that training was not affected by game factors such as art style and narrative suggests that the quality of training can trump aesthetic factors. Future research should investigate whether this holds true for other types of training.

**References**

Aldrich, C. (2005). *Learning by doing: A comprehensive guide to simulations, computer games, and pedagogy in e-learning and other educational experiences*. San Francisco, CA: Pfeiffer.

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

Arnott, D. (2006). Cognitive biases and decision support systems development: A design science approach. *Information Systems Journal*, *16*, 55–78. doi:10.1111/j.1365-2575.2006.00208.x

Bacon, D. (2004). The contributions of reliability and pre-tests to effective assessment. *Practical Assessment, Research and Evaluation*, *9*(3). Retrieved from http://pareonline.net/getvn.asp?v=9&n=3

Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and co-presence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, *15*, 359–372. doi:10.1162/pres.15.4.359

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn*. Washington, DC: National Academy Press.

Clark, D., Tanner-Smith, E., & Killingsworth, S. (2014). *Digital games, design and learning: A systematic review and meta-analysis*. Menlo Park, CA: SRI International.

Clegg, B. A., Martey, R. M., Stromer-Galley, J., Kenski, K., Saulnier, T., Folkestad, . . . Strzalkowski, T. (2014). Game-based training to mitigate three forms of cognitive bias. *Proceedings of Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, 14180, 1–12.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York, NY: Academic Press.

Conle, C. (2003). An anatomy of narrative curricula. *Educational Researcher*, *32*(3), 3–15. doi:10.3102/0013189X032003003

Cook, M. B., & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Human Factors*, *50*, 745–754. doi:10.1518/001872008X354183

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 2: Impediments to and strategies for change. *BMJ Quality and Safety, 22*, ii65–ii72. doi:10.1136/bmjqs-2012-001713

Dale, R., Kehoe, C. E., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory and Cognition*, *35*, 15–28. doi:10.3758/BF03195938

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

Dettori, G., & Paiva, A. (2009). Narrative learning in technology-enhanced environments: An introduction to narrative learning environments. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning, part 1* (pp. 55–69). New York, NY: Springer.

Dickey, M. D. (2006). Game design narrative for learning: Appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology and Research Development*, *54*, 245–263. doi:10.1007/s11423-006-8806-y

Eisner, E. W. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Upper Saddle River, NJ: Prentice Hall.

Fischer, P., & Greitemeyer, T. (2010). A new look at selective-exposure effects: An integrative model. *Current Directions in Psychological Science*, *19*, 384–389. doi:10.1177/0963721410391246

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

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. London, UK: Palgrave Macmillan.

Gerrig, R. J. (1993). *Experiencing narrative worlds*. New Haven, CT: Yale University Press.

Gilbert, D. T., & Malone, P. S. (1991). The correspondence bias. *Psychological Bulletin*, *1117*, 21–38. doi:10.1037/0033-2909.117.1.21

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *29*, 207–219. doi:10.1111/j.1365-2729.2012.00489.x

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, *14*, 311–327. doi:10.1111/j.1468-2885.2004.tb00317.x

Hinyard, L. J., & Kreuter, M. W. (2007). Using narrative communication as a tool for health behavior change: A conceptual, theoretical, and empirical overview. *Health Education and Behavior*, *34*, 777–792. doi:10.1177/1090198106291963

Holding, D. H. (1991). Transfer of training. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 93–125). New York, NY: John Wiley.

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling*, *61*, 173–190. doi:10.1016/j.pec.2005.05.004

Hughes, D. (2014). *The design and evaluation of a video game to help train perspective-taking and empathy in children with autism spectrum disorder* (Unpublished doctoral dissertation). University of Central Florida, Orlando, FL.

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

Kirriemuir, J., & McFarlane, A. (2008). *Literature review in games and learning* (No. 8). Futurelab Series. Berkshire, UK: National Foundation for Educational Research in England and Wales. Retrieved from http://archive.futurelab.org.uk/resources/documents/lit_reviews/Games_Review.pdf

Koenig, A. (2007, July). *Exploring narrative as a cognitive aid in educational video games*. Paper presented at the 13th International Conference on Artificial Intelligence in Education, Marina Del Rey, CA.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.

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

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal*, *30*, 195–232.

Li, M. C., & Tsai, C. C. (2013). Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*, *22*, 877–898. doi:10.1007/s10956-013-9436-x

Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52. doi:10.1207/S15326985EP3801_6

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York, NY: Penguin Press.

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008). Story-based learning: The impact of narrative on learning experiences and outcomes. In *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems* (pp. 530–539). Berlin, Germany: Springer.

Michael, D. R., & Chen, S. (2006). *Serious games: Games that educate, train, and inform*. Stamford, CT: Cengage Learning.

Mott, B. W., Callaway, C. B., Zettlemoyer, L. S., Lee, S. Y., & Lester, J. C. (1999). Towards narrative-centered learning environments. In M. Mateas & P. Sengers (Eds.), *Narrative intelligence: Papers from the 1999 fall symposium* (pp. 78–82). Menlo Park, CA: American Association for Artificial Intelligence.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. doi:10.1080/00335557743000053

Nowak, K. L., Hamilton, M. A., & Hammond, C. C. (2008, October). *The effect of viewer perceptions of avatar anthropomorphism and realism on judgments of credibility and homophily, and avatar choice*. Paper presented at PRESENCE 2008: 11th Annual International Workshop on Presence, Padova, Italy.

Peng, W., Lee, M., & Heeter, C. (2010). The effects of a serious game on role-taking and willingness to help. *Journal of Communication*, *60*, 723–742. doi:10.1111/j.1460-2466.2010.01511.x

Peng, W., & Liu, M. (2008). An overview of using electronic games for health purposes. In R. E. Ferdig (Ed.), *Handbook of research on effective gaming in education* (pp. 388–401). Hershey, PA: IGI Global.

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, *11*, 37–43. doi:10.1111/j.1460-2466.2010.01511.x

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*, 369–381. doi:10.1177/0146167202286008

Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent group designs. In T. D. Cook & D. T. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings* (pp. 147–205). Boston, MA: Houghton Mifflin.

Ritterfeld, U., Cody, M., & Vorderer, P. (2010). *Serious games: Mechanisms and effects*. London, UK: Routledge.

Rowe, J., McQuiggan, S., Mott, B., & Lester, J. (2007, July). *Motivation in narrative-centered learning environments*. Paper presented at the AIED 2007 Workshop on Narrative Learning Environments, Marina del Rey, CA.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432–1463. doi:10.1037/a0037559

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

Sanna, L. J., Schwarz, N., & Small, E. M. (2002). Accessibility experiences and the hindsight bias: I knew it all along versus it could never have happened. *Memory and Cognition*, *30*, 1288–1296. doi:10.3758/BF03213410

Sax, G. (1997). *Principles of educational and psychological measurement and evaluation* (4th ed.). Belmont, CA: Wadsworth.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217. doi:10.1111/j.1467-9280.1992.tb00029.x

Shaw, A. (2014). *Gaming at the edge: Sexuality and gender at the margins of gamer culture*. Minneapolis, MN: University of Minnesota Press.

Simpson, T., Camfield, D., Pipingas, A., Macpherson, H., & Stough, C. (2012). Improved processing speed: Online computer-based cognitive training in older adults. *Educational Gerontology: An International Journal*, *38*, 445–458. doi:03601277.2011.559858

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, *64*, 489–528. doi:10.1111/j.1744-6570.2011.01190.x

Stewart, T. L., Latu, I. M., Kawakami, K., & Myers, A. C. (2010). Consider the situation: Reducing automatic stereotyping through situational attribution training. *Journal of Experimental Social Psychology*, *46*, 221–225. doi:10.1016/j.jesp.2009.09.004

Stokes, S. (2002). Visual literacy in teaching and learning: A literature perspective. *Electronic Journal for the Integration of Technology in Education*, *1*, 10–19. Retrieved from http://ejite.isu.edu/Volume1No1/Stokes.html

Sun, Y., Pan, Z., & Shen, L. (2008). Understanding the third-person perception: Evidence from a meta-analysis. *Journal of Communication*, *58*, 280–300. doi:10.1111/j.1460-2466.2008.00385

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*, 229–243.

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, *103*, 506–519. doi:10.1037/a0028857

Wolf, M. J. P. (2003). Abstraction in the video game. In M. J. P. Wolf & B. Perron (Eds.), *The video game theory reader* (pp. 47–65). New York, NY: Routledge.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*, 249–265. doi:10.1037/a0031311

Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the e-junior project. *Computers and Education*, *55*, 178–187. doi:10.1016/j.compedu.2010.01.003