Testing the Power of Game Lessons: The Effects of Art and Narrative on Reducing Cognitive Biases

Rosa Mikeal Martey
Colorado State University
Department of Journalism and Technical Communication
Campus Delivery 1785
Fort Collins, CO 80523-1785
rosa.martey@colostate.edu

Adrienne Shaw
Temple University
School of Media and Communication
Department of Media Studies and Production
2020 N. 13th St.
Philadelphia, PA 19104
adrienne.shaw@temple.edu

Jennifer Stromer-Galley
Syracuse University
School of Information Studies
220 Hinds Hall
Syracuse, NY 13244
jstromer@syr.edu

Kate Kenski
University of Arizona
Department of Communication
1103 E. University Blvd
P.O. Box 210025
Tucson, AZ 85721
kkenski@email.arizona.edu

Benjamin A. Clegg
Colorado State University
Department of Psychology
1876 Campus Delivery
Fort Collins, CO 80523-1876
Benjamin.clegg@colostate.edu

Proceedings of DiGRA 2014: <Verb that ends in 'ing'> the <noun> of Game <plural noun>.

James E. Folkestad
Colorado State University
School of Education
244 Education Building
Fort Collins, CO 80523-1876
James.folkestad@colostate.edu

Emilie T. Saulnier
1st Playable Productions
5 Third St. Suite 300
Troy, NY 12180
tobi@1stplayable.com

Tomek Strzalkowski
SUNY-Albany
ILS Institute
1400 Washington Ave
Albany, NY 12222
tomek@albany.edu

ABSTRACT

Educational games have proliferated, but questions remain about the effectiveness at teaching both in the short- and long-term. Also unclear is whether particular game features have positive effects on learning. To examine these issues, this paper describes a controlled experiment using an educational game that was professionally developed to teach about cognitive biases in decision making (Fundamental Attribution Error, Confirmation Bias, and Bias Blind Spot). This experiment examined the effects of game art and narrative on learning and compared the game conditions to a training video. Effects were measured immediately after the stimuli were given and then again eight weeks later. Results indicate that the educational game outperforms the training video immediately after exposure and that there are significant retention effects. Art and narrative were not significantly related to learning with the exception that minimal art game had a significant positive relationship with mitigating Bias Blind Spot at immediate post-test.

Keywords

cognitive biases, educational games, game characteristics, narrative, art

INTRODUCTION

Cognitive biases, such as confirmation bias or fundamental attribution error, are a human tendency to use shortcuts in information processing and decision-making (Kahneman, 2011). A body of scholarship has examined the triggers and manifestations of cognitive biases, but fewer studies have found effective ways to teach people how to overcome this fact of our brains in order to help us make better decisions.

Educational or "serious" games have grown both as an area of product development for game companies and as a subject of research to ascertain effectiveness. Of particular interest in the scholarship on educational games is their ability to teach people more successfully, with learning outcomes that exist over time, and especially when compared to more traditional teaching methods, such as training videos or lecture materials.

The purpose of this project is to ascertain the effectiveness of an educational game to teach about cognitive biases in decision-making and to teach players to mitigate those biases. We also seek to understand whether there are particular aspects of game design that might more effectively impart learning. This project tests whether the complexity level of narrative and art style have effects on learning immediately after playing an educational game as well as eight weeks later.

Below, we briefly describe the literature on educational games, cognitive biases, and the game components of art and narrative. We then describe our research methods, which involved building and testing an educational game using a 2x2 full factorial design experiment manipulating levels of art and narrative complexity, plus a training video for comparison. Results indicate that the game significantly outperformed the comparison video on measures of cognitive biases on a survey immediately after the stimulus, and that the significant effects were evident eight weeks later. The art and narrative conditions produced a complex picture of design effects, with minimal art having a significant improvement on learning compared with detailed art on one of the cognitive biases examined.

COGNITIVE BIASES AND SERIOUS GAMES

Research on cognitive biases demonstrates that biases function as useful heuristics when sorting through large amounts of information, when certain information primes them to focus on specific types of information, or when they are asked to make quick decisions/analyses. This has been demonstrated for confirmation bias (Fischer & Greitemeyer, 2008) and fundamental attribution error (Jones & Harris, 1967; Tetlock, 1985). Confirmation bias is the tendency to seek out or favor information that supports pre-existing beliefs (Kahneman & Tversky, 1973). Fundamental attribution error is a related bias that is the tendency for people to attribute personality-based explanations for the behaviors they observe in others (Jones & Harris, 1967). People are largely unaware of their reliance on these biases and thus have a *bias blind spot* (Pronin, Lin & Ross, 2002).

Getting people to recognize the shortcomings of these heuristics is a difficult task. Research on bias mitigation techniques generally argues that people must be taught to question reflexively their initial evaluations of scenarios in order to account more thoroughly for alternative interpretations and unaccounted for variables (Stewart, Latu, Kawakami, & Meyers, 2010). Digital games offer one tool for training these techniques (Rahford, 2009). Digital games have a well-documented potential for offering new and innovative learning environments. As Malone (1981), Rieber (1996), Gee (2003), and Aldrich (2005) identify, good games mirror effective learning models.

More research is needed to determine how games might be used for bias mitigation training. Although there have been efforts to reduce bias in specific decision-making contexts, these are typically framed at the point of decision-making (Larrick, 2004), or are based on tools that can directly influence the processing of information (e.g. decision support systems) within the actual task environment (Arnott, 2006). Though reducing individuals' overall tendencies toward bias through training was proposed by Fischhoff (1982), implementing such training has had limited success (Sanna, Shorts, & Small, 2002). In addition, though some bias elicitation and training techniques have been found to be reliable in non-game settings (Dale, Kehoe, & Spivey, 2007), the use of these techniques in a game environment has not been evaluated. It has been argued that intrinsic motivation generates more learning outcomes (Deci & Ryan, 1985) and in particular that effective games are intrinsically motivating to players (McGonigal, 2011; Salen & Zimmerman, 2003). Further research to examine how a digital game compares with other teaching techniques, such as an educational video, are needed to better illuminate the effectiveness of the delivery mechanism for teaching. This study, thus, asks six interrelated research questions:

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

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

In addition to testing these broader questions, we also are interested in the impact of specific game variables on learning outcomes. Although many researchers have discussed the importance of narrative in game spaces, few have analyzed how differences in narrative richness affect learning (Dickey, 2006). Use of a story format has been found to increase learning and engagement (Hinyard & Kreuter, 2007; Dettori & Paiva, 2009). Specifically, researchers suggest that rich narratives promote reflexivity (Conle, 2003; Eisner, 1998), lead to exploratory learning (Mott et al., 1999), provide motivating learning scenarios (Rowe et al., 2007), and lead to transportation and engagement (Gerrig, 1993; Green et al., 2004). Moreover, kinds of narratives in games are also diverse including: evoked narratives, enacted narratives, embedded narratives, and emergent narratives (Jenkins, 2004). In this experiment we wanted to see if the level of narrative would affect learning outcomes.

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

Narratives cannot be experienced in isolation, however. As Jenkins (2004) argues the sensory environment of games can tell a story just as much as cut-scenes and dialog. Indeed in a previous experiment we found that, even if a game was designed and constructed within little or no narrative, participants constructed narratives for a game on their own (authors). There was some ambivalence from those interviewees over the need for a story, and it is unclear what sort of story would really make the content more coherent to players. To strengthen the potential impact of the narrative, then, we also manipulated the level visual abstraction. Some scholars suggest that visual abstraction leads players to be *more* engaged in games (Wolf & Wolf, 2003). Other researchers argue that realism enhances feelings of co-presence (Bailenson et al., 2006) and that realism affects avatar credibility (Nowak et al., 2008). To that end, we ask:

RQ4: Does **detailed art** produce better bias reduction between pre- and post-tests than minimal art?

In addition, we sought to investigate whether learning effects lasted. Thus:

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

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

By studying significant effects of conditions as well as participants' qualitative assessments of conditions, we argue that we are better able to parse how certain game variables come to be important in Serious Games.

METHODS

Stimuli

To examine whether detailed art or a rich narrative have an impact on bias mitigation we built an educational game working with a professional educational game company. We used a 2 x 2 + comparison video full factorial experiment, Rich (RN) versus Light Narrative (LN) and Detailed (DA) versus Minimal Art (MA), to examine bias mitigation outcomes. Note that this study does not use a traditional control condition, but rather a professionally-produced video that trains in the same content but with different format and language than the game tested. This was done in order to compare the game outcomes with those from material more likely to be used in a training or teaching setting.

The game is a flash-based puzzle game in which players navigate an avatar through a series of rooms containing puzzles that teach about the three biases and how to mitigate them. Players also received infographics before each room to define and describe biases and mitigation strategies, and short interactive quizzes followed each room to further reinforce each lesson. To minimize the potential effect of disidentification in the game the player avatar was as gender and racially neutral as possible (Muñoz, 1999; Shaw, 2010). It was dressed in a gender-neutral training suit with a helmet to obscure face and hair. A help system was available that provided additional instructions for players who were stuck, and hints occasionally appeared at the bottom of the screen to remind players of the task or to provide key information needed to make progress in the room. Our experimental stimuli were four game conditions that varied art (detailed vs. minimal) and narrative complexity (rich vs. light).

Across the two art conditions the perspective and functionality of the game were kept the same, but the DA condition was full-color, with rich texture and shading and realistic detailing. The MA condition was largely monochrome with minimal shading and almost no textures (see Figure 1). A few additional detail objects were added to the backgrounds for the DA condition, but no interactivity differences were created between the two variables.

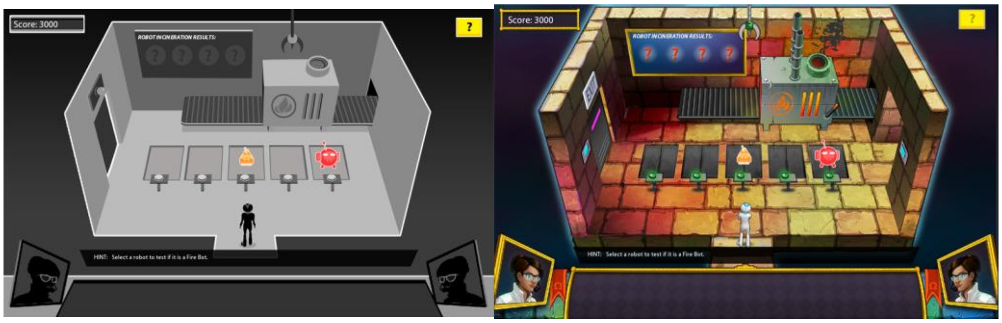

Figure 1: Minimal and Detailed Art

For each of the art conditions, a LN version of the game was built. In the LN condition the player was only told he or she was in a training center to learn about cognitive biases. The RN version positioned the player as the child of a bias reduction expert who once worked with the CYCLE Center owner, Dr. Ohm. In the game narrative, the player infiltrates the training 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.

Much of the RN story was told in an introduction, in text added to transition rooms, and in two possible conclusions a player could get depending on a choice made in the final room. Puzzle room text was almost exactly identical across the narrative conditions to maintain equivalent content delivery. VO was adjusted for the delivery of specific plot lines in the HN condition to be more dramatic and character-driven. Aside from the introduction and conclusions of the games, which used version-specific splash screens, there was no art generated specifically for the narrative.

The training video was professionally written, acted, and produced and provided the same core information as the video game with regard to the definitions and general bias mitigation strategies, but did so by providing a series of scenarios in which people exhibited cognitive biases. Those scenarios were then dissected by a "professor" who explained the interaction, the causes of bias and ways to mitigate the bias. The topics of this video were the same as the game, but the language, examples, and teaching approach were different in order to create effective training in video format.

Study Procedures

Participants were recruited from college classes and psychology subject pools at three universities in the United States in the Fall semester of 2012. Participants were then 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. Items included university, internet/video game experience, FAE, CB, BBS, personal bias awareness, social conformity, Need for Cognition, gender roles, Big Five personality assessment, and demographics. Participants were then randomly assigned to one of the five study conditions. Immediately after completing the game/video, they answered a 20-minute post-session questionnaire. Items included engagement, usability (ease, attention, stimulation, likability), mitigation and knowledge of the three biases, and some manipulation check questions. A final questionnaire was administered online approximately eight weeks later that included transportability (personality trait), modified engagement and usability questions, learning perceptions, and mitigation and knowledge of the three biases.

Study Participants

We had 480 subjects included in our final dataset, after excluding 16 participants for errors in data collection. One hundred eighty four people participated in the retention study (for a 40% response rate), which participants accessed remotely via the internet.

The majority of our sample was female (65% female; 35% male), 79% were between the ages of 18 and 20. Slightly over half (54%) of participants reported they have an intermediate level computer skills, 23% advanced, 10% expert, and 13% novice. Most participants did not consider themselves “gamers” (84%). Twenty % reported playing games less than once every few weeks, 19% every few weeks, 17% 1-2 days a week, 14% 3-5 days a week, 13% about once a day, and 17% several times a day.

Measures

We developed three principal mitigation scales to measure the cognitive biases that were the focus of study: fundamental attribution error (FAE), confirmation bias (CB), and bias blind spot (BBS). The bias mitigation measures were developed using extensive pre-testing including examining them for testing effects that found no significant pre-post differences when the measures were administered before and after a filler activity of doing math problems.[1]

Fundamental Attribution Error (FAE)

This measure presented participants with ten brief scenarios such as “One of your peers receives an 'A' in a course that has a reputation for being hard. The best explanation for this student’s grade is that the student is smart. To what extent do you agree or disagree with this assessment?” Scenarios were based in range of topics areas such as donating money, romantic relationships, and driving. Answers used sliders on a scale of 1 (*disagree*) to 7 (*agree*), and were scored by taking the mean of all 10 ratings and transforming them to a -100 to +100 scale.

Confirmation Bias (CB)

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**,** participants are presented with seven scenarios such as “You are considering taking a trip to the country Calzycoah. You don’t speak the language, but you will not have to pay for airfare to this vacation spot.” Then some scenarios provided a hypothesis, others asked participants to decide between two possible hypotheses. Participants then rated six pieces of evidence 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*). Items were calculated by subtracting the mean of the confirming items from the mean of the disconfirming items and adjusting them to a -100 to +100 scale.

Bias Blind Spot (BBS)

BBS was measured with a variant of the method employed by Pronin et al. (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 (IS), the tendency to rate oneself as above average on these types of dimensions. They were then asked, "To what extent do you believe that you showed this tendency when you rated your [intelligence] on a previous question?" This was followed with: "To what extent do you believe that the average student from your university would show this tendency if he or she rated his or her [intelligence]?" Answers are on sliders from 1 (*not at all*) to 9 (*very much*) and were scored by subtracting ratings of others from ratings of the self, taking the mean, and adjusting to a -100 to +100 scale.

RESULTS

Bias Mitigation, Art, Narrative, and Retention

Overall the experiment results showed that training with all formats of materials reduced cognitive biases, with a superior training performance from the game. Table 1 shows the means and standard deviations for each condition of the combined biases measure.

	Pre-test			Post-test			8-week retention		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
LN/MA	99	19.91	13.47	99	2.44	16.49	30	10.46	16.40
LN/DA	80	22.37	13.85	93	5.11	15.96	36	11.42	13.23
RN/MA	95	19.70	12.92	95	-0.61	16.21	40	6.70	17.81
RN/DA	104	21.04	13.08	104	3.67	13.39	38	8.17	13.53
Video	89	18.65	11.98	89	14.82	16.63	40	17.36	11.79

Table 1: Immediate and Delayed Bias Mitigation Descriptive Statistics Combined Biases

To examine whether or not a video game can train cognitive biases (RQ1), we used a repeated measures analysis of variance (DM MANOVA) comparing the three bias measures at pre-test with immediate post-test for the game condition only (N = 391). To examine RQ2 through RQ4, we used a pair of analyses: first an ANOVA featuring the repeated measures bias score from pre-test and post-test by the 5 conditions; and second, an examination of post-test only performance by condition using planned linear contrasts to explore the effects of game play versus the video (LN/MA & LN/DA & RN/MA & RN/DA vs. video), differences between the art style conditions (LN/MA & RN/MA vs. LN/DA & RN/DA), and differences between the narrative style conditions (LN/MA & LN/DA vs. RN/MA & RN/DA). The same approach is used to examine RQ5 and RQ6 comparing pre-test to the 8-week post-test. Here, we present these results for all biases combined and for each bias individually.

Bias Reduction at Immediate Post-Test

Overall, across all baises, we found a significant effect of training from pre- to immediate post-test (RQ1) for the game conditions (Wilks' lambda = .48, F(3,390) = 426.50, $p < .0005$), but no difference between conditions (F(1,387) = .70, $p > .05$). The video condition was not included in this analysis.

Examining FAE for all conditions, we found a change from training (see Figure 2), with a main effect of pre-post (F(1,475) = 393.18, $p < .0005$, MSE = 726), a main effect of condition (F(4,475) = 4.28, $p < .05$, MSE = 1604), and an interaction between pre-post and condition (F(4,475)=8.47, $p<.0005$). Examining the post-test performance using the planned linear contrasts (RQ3 and RQ4) showed an advantage in FAE training for game play versus the video (RQ2): all game conditions vs. video; t(475) = -5.26, $p < .0005$. There was no difference between the game art conditions (LN/MA & RN/MA vs. LN/DA & RN/DA; t(475) = -1.41, $p > .05$), and no difference between the game narrative conditions (LN/MA & LN/DA vs. RN/MA & RN/DA; t(475) = 0.48, $p > .05$).

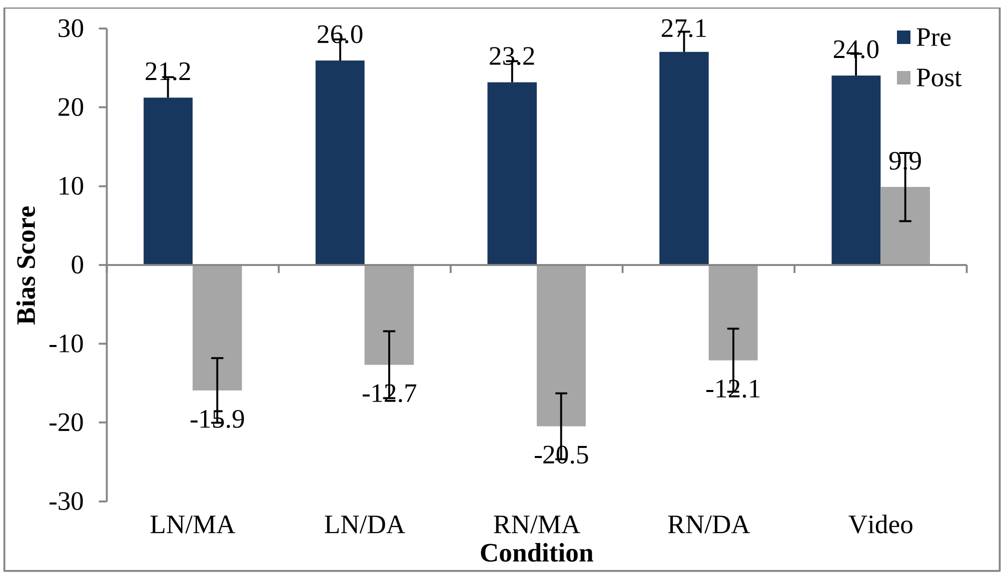

Figure 2: Fundamental Attribution Error by Condition for Pre-test and Post-test

CB also showed a change from training (see Figure 3), with a main effect of pre-post (F(1,475) = 107.41, $p < .0005$, MSE = 187), a main effect of condition (F(4,475) = 2.47, $p < .05$, MSE = 550), and an interaction between pre-post and condition (F(4,475)=8.11, $p<.0005$). Planned linear contrasts showed an advantage for game play versus the video (RQ2): all game conditions vs. video; t(475) = -4.61, $p < .0005$. There was no difference between the game art conditions (LN/MA & RN/MA vs. LN/DA & RN/DA; t(475) = -0.19, $p > .05$), and no difference between the game narrative conditions (LN/MA & LN/DA vs. RN/MA & RN/DA; t(475) = 1.66, $p = .09$).

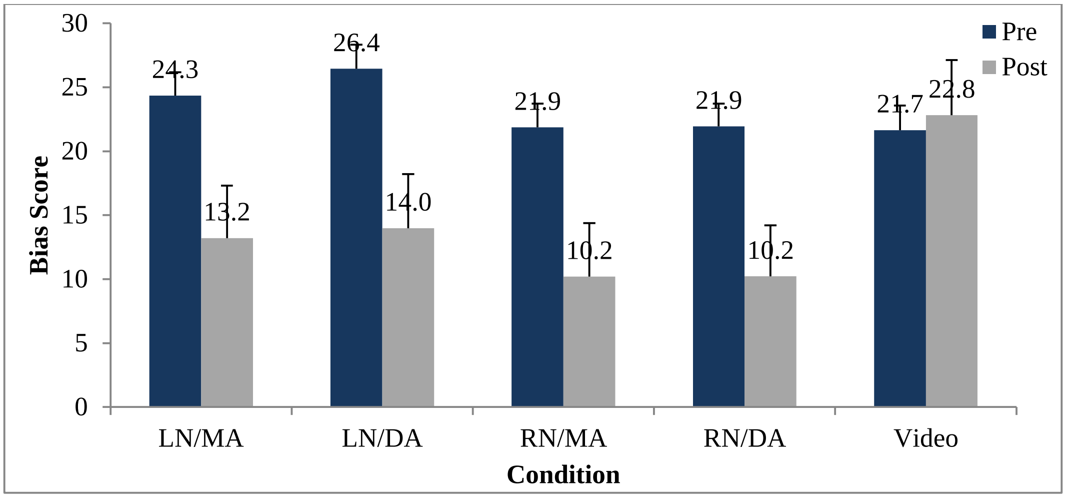

Figure 3: Confirmation Bias by Condition for Pre-test and Post-test

BBS showed no overall change from training (see Figure 4), with the main effect of pre-post marginally non-significant (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). Planned comparisons revealed, however, that one game condition, RN/MA, showed significant reduction in bias from pre- to post-test (t(94) = 2.67, *p* < .01), but that there were no significant reductions in bias for the LN/MA condition (t(98) = 1.75, *p* =.08), for LN/DA (t(92) = 0.28, *p* > .05), or for RN/DA (t(103) = 0.49, *p* > .05), and no training benefit from the video condition (t(88) = -0.72, *p* > .05).

Post-hoc planned linear contrasts showed no overall advantage for game play versus the video (all game conditions vs. video; t(475) = -0.19, *p* >.05), a difference between the game art conditions (LN/MA & RN/MA vs. LN/DA & RN/DA; t(475) = -2.10, *p <.05*), no difference between the game narrative conditions (LN/MA & LN/DA vs. RN/MA & RN/DA; t(475) = 0.67, *p* >.05).

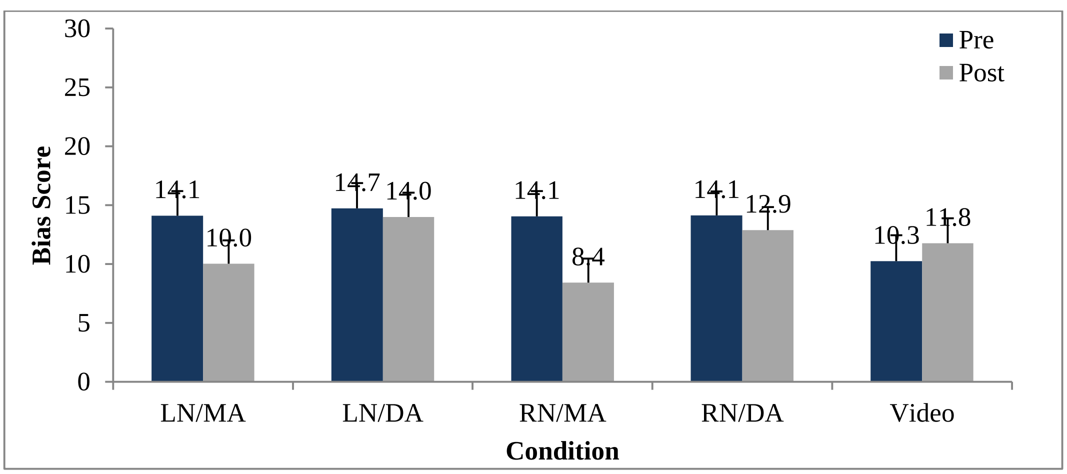

Figure 4: Bias Blind Spot by Condition for Pre-test and Post-test

Results at 8-week Retention Test

Overall the experiment results showed that training with all formats of materials reduced cognitive biases at 8-weeks (RQ5). A repeated measures analysis of variance (DM MANOVA) comparing pre-test bias to the 8-week retention bias across the three types of bias as dependent variables showed a significant pre to retention effect of training for the game conditions (Wilks' lambda = .64, F(1,143) = 80 p < .0005), but no difference between conditions (F(1,140) = .357, p > .05). The video condition was not included in this analysis. Here we present analyses for each bias.

Examining FAE across all conditions showed a change at 8-weeks from training (see Figure 6), with a main effect of pre to retention bias levels (F(1,179) = 51.06, p < .0005, MSE = 557), no main effect of condition (F(4,179) <1, MSE = 1350), and an interaction between pre-retention and condition (F(4,179)=3.05, p<.05). 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 conditions (LN/MA & RN/MA vs. LN/DA & RN/DA; t(179) = 0.49, p >*.05*), and no difference between the game narrative conditions (LN/MA & LN/DA vs. RN/MA & RN/DA; t(179) = 0.82, p >*.05*).

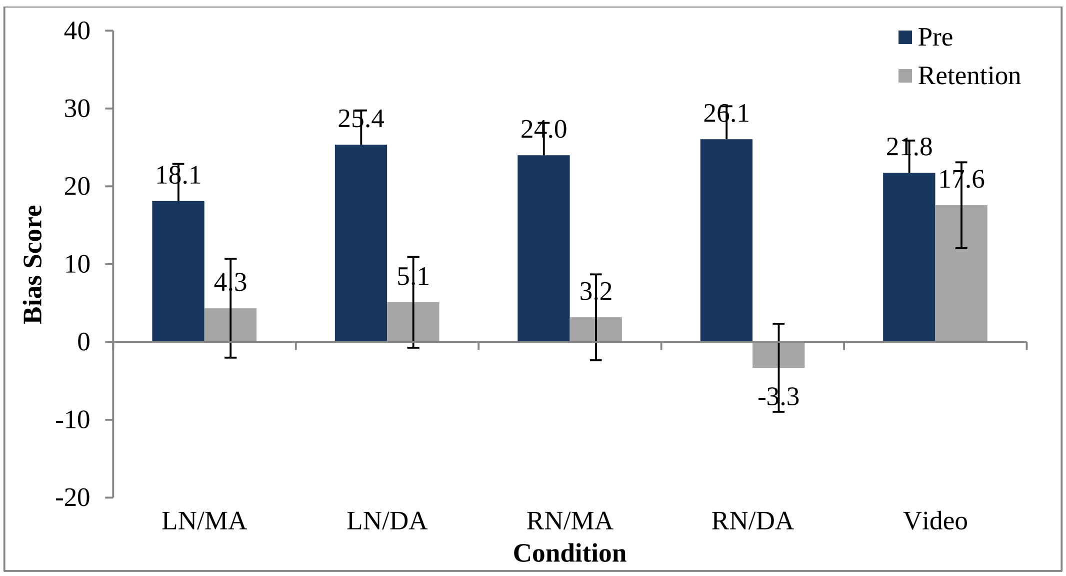

Figure 6: Fundamental Attribution Error by Condition for Pre-test and Retention test

CB showed a change from training (see Figure 7), with a significant main effect of pre-test to retention-test (F(1,179) = 50.59, p < .0005, MSE = 227), 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 conditions (LN/MA & RN/MA vs. LN/DA & RN/DA; t(179) = 0.25, p >*.05*), no difference between the game narrative conditions (LN/MA & LN/DA vs. RN/MA & RN/DA; t(179) = 0.17, p >*.05*).

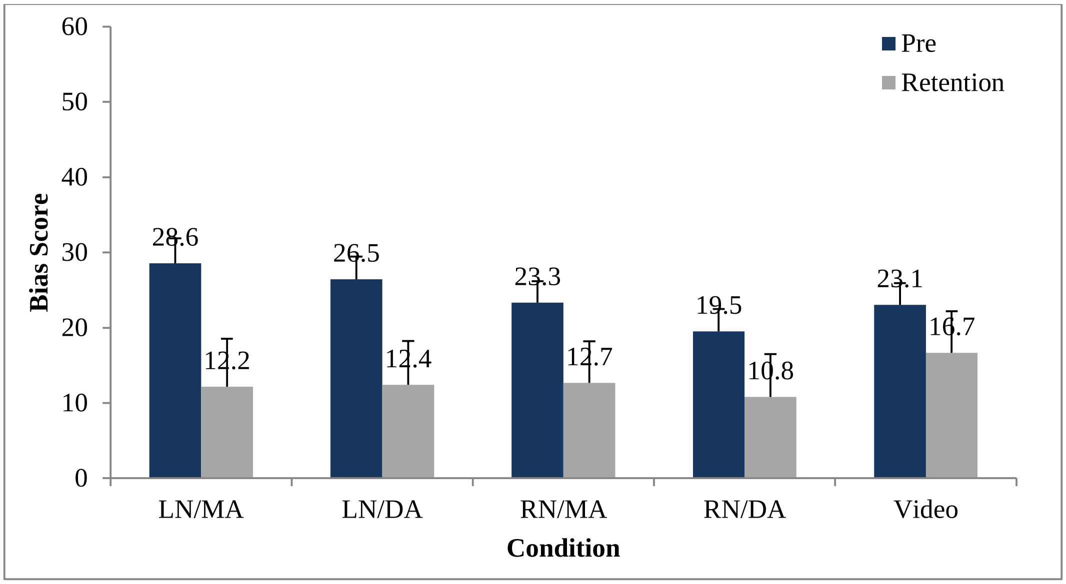

Figure 7: Fundamental Attribution Error by Condition for Pre-test and Retention test

BBS showed a change from training (See Figure 8), 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 (F(4,179)<1). Examining the retention test performance using the post-hoc planned linear contrasts showed no significant advantage for all game play versus the video (all game conditions vs. control video; t(179) = 1.43, $p > .05$), no difference between the game art conditions (LN/MA & RN/MA vs. LN/DA & RN/DA; t(179) = -0.27, $p > .05$), no difference between the game narrative conditions (LN/MA & LN/DA vs. RN/MA & RN/DA; t(179) = 0.46, $p > .05$).

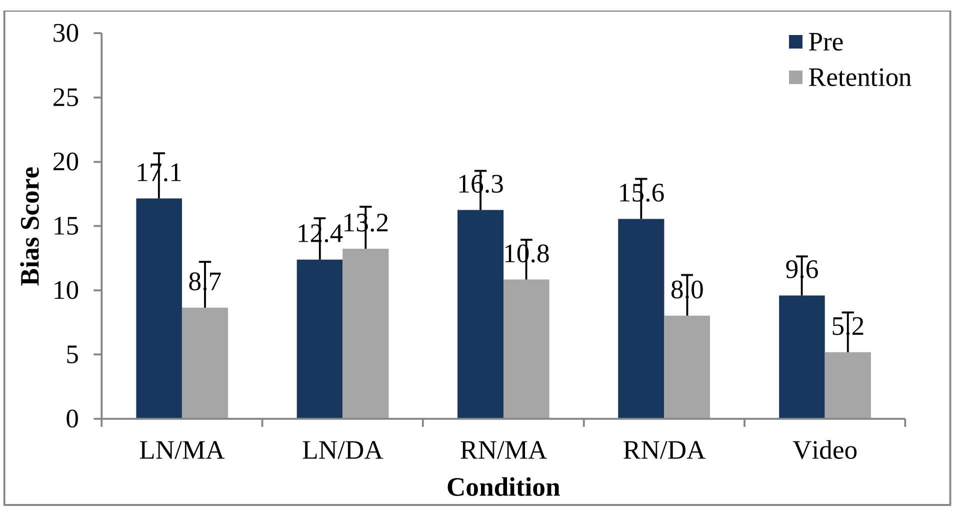

Figure 8: Bias Blind Spot by Condition for Pre-test and Retention test

DISCUSSION

The results of the experiment suggest that an educational game that teaches people how to mitigate them generally performed better than a training video immediately after game play, and that the effects lasted eight weeks later. Although we experimented with art and narrative and examined the interactions of art and narrative, we did not find significant results, except in the relationship between BBS and minimal art at immediate post-test. That is, one minimal art condition (with rich narrative) was the only condition in which there was a statistically significant drop in BBS from pre- to immeidate post-test, but not retention.

Bias mitigation is more challenging for some biases than others. BBS was only reduced in one of the five conditions. As Pronin and others have found, BBS can actually *increase* among those with greater knowledge of biases or 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 because an increased *awareness* of biases leading to increased confidence that biases can be completely avoided. Indeed, although BBS showed little to no improvement at immediate post-test, it did show improvement at the 8-week retention test, suggesting that distance from the training was actually better for BBS reduction. This may be why our results on this measure are lower than others.

The greatest improvement on mitigation is in FAE. Our training in FAE also seems to be the clearest, most memorable for participants. This might be due to the game's clear mitigation strategy, which participants explain as, "look for environment and not just the person." It is important to note that although the game trained significantly better, the video also trained participants to a 59% reduction, suggesting that overall, FAE may be easier to train against than the other biases.

Achieving reduction in CB is challenging. The complexity of this concept and its application make expressing the definition and mitigation strategies for CB more difficult. We also suggest that CB manifests distinctly in different types of decisions or contexts. The most substantial challenge seems to be helping players understand how to translate CB mitigation to new contexts. We currently have several different settings for CB training, including real-world scenarios presented in quiz questions. These seem to help, as answering the quiz questions correctly corresponds with reductions in biases on the post-test survey.

More generally, we feel CB is easiest to understand when it is first *experienced*. That is, training could be more effective when a player makes a decision based on CB and then understands how this decision was subject to bias. We address this by starting with applied lessons and then unpacking the steps required to identify confirming and disconfirming information, then applying a system of checking both in new scenarios. However, such lessons are difficult to impart across different types of people with different play styles, interests, and backgrounds. Some participants still expressed confusion over the concept.

One challenge of this experiment was in the design of the bias measures. As there are not effective and widely used batteries for all three biases, we had to develop our own. One limitation of this work is that these items are not widely validated. We have worked to address these issues by conducting analyses to compare our BBS measures against Pronin et al.'s (2002) original work to determine if our measures capture the same phenomenon she identifies. Second, we have worked to determine conceptually and with data from additional studies what the most meaningful "no bias" point should be in our FAE and BBS scales. Although differences between assessing individuals and the environment for FAE make theoretical sense, we also want to understand how these assessments compare to other possible explanations. For example, when considering why a student failed an exam, how does the explanation "the person is lazy" compare to other possible explanations such as "the exam is difficult" or "the person studied the wrong material"? We have investigated how answers to our items relate to other plausible explanations for the scenarios in our FAE questions as part of an on-going analysis. For BBS, a key aspect of a non-biased zero point is the reference group for self and other assessments.

Interviews we conducted with 19 participants (selected to balance gender, race, and game condition) help further illuminate the art and narrative as experienced by players. Specifically, when we compared art conditions, we saw differences in players' overall assessment of the game by art condition. In general, players who were in the minimal art condition liked the game more than those in the detailed art condition. Some participants wanted more color in the game, but few of those with the highly colorful detailed art mentioned color as a core reason for the game's appeal.

The art condition also seemed to shape interpretations and expectations of the narrative condition and vice versa. Those in the rich narrative/detailed art condition expressed more confusion about the story and were not as responsive to it as those in the rich narrative/minimal art condition. Those in rich narrative/detailed art condition were also more critical of the lack of integration of the story and the training, as well as the game overall. All of them properly identified basic elements of the story, but all of them either stopped paying attention to the story or felt the story only existed in the background and was not related to what they did in the game. As one interviewee in the rich narrative/detailed art condition stated: "You know, I think you need to make a choice between having a story and not having a story. I mean, if it was just a training exercise where you're trying to learn about biases—I mean, it was trying to do that, and at the same time, you had this evil doctor [Dr. Ohm] that you're sneaking into his lab to do it with, and it conflicted and it didn't make a whole lot of sense." Those in the rich narrative/minimal art condition also identified the story and generally felt the story was good enough for the game. One said: "I thought the story was good. It made me laugh." Although participants in the light narrative/minimal art condition did not see the game as having a story, many of the light narrative/detailed art interviewees did identify some basic narrative elements.

It also seems that narrative condition shapes how players rate the art. All of those in minimal narrative liked the art (regardless of condition). One minimal art interviewee said the lack of detail helped him focus on the puzzles in the rooms: "I guess it kind of helped me focus on the stuff that I needed to click on or answer or whatever." A participant in the detailed art condition said, "I liked the game and the different elements… the lighting and the brick." The interviewees in the rich narrative conditions were not as positive about the art, comparing the game to commercial games and expressing a desire to see higher production values. Those in the rich narrative condition who liked the art pointed to elements like the brightness. The combination of detailed art and rich narrative may have increased players' expectations of the game.

CONCLUSION

The burgeoning research on educational games suggests they might more effectively impart learning on a variety of subjects (Aldrich, 2005; Gee, 2003; Malone, 1981; Rieber, 1996). In our experiment, our game did indeed outperform a training video of similar content in helping participants learn to mitigate biases. The training was also retained 8 weeks later.

Prior research on the level of art detail provides a mixed picture on whether more abstract art leads players to be *more* engaged in games (Wolf, 2003), which lead to greater learning, or if more realistic art enhances feelings of co-presence (Bailenson et al., 2006), which might improve learning. Results from our experiment suggest that minimalistic art outperforms detailed art on one of the cognitive biases, BBS. However, given the overall poor performance of BBS training and the lack of interaction effects in the model examining BBS, it is unclear if this result is robust, as this effect disappears at the 8-week retention-test. More research is needed to further parse out the role of art on learning in educational games. What our results suggests is that there may be greater learning with minimalistic art because the reduced cognitive load of processing detail and color in the visual channel may improve overall short-term and long-term memory of the material being taught.

The literature also does not provide clarity about the role of narrative on learning. Scholars have tended to assume that narrative matters in game spaces, but the question has not been systematically tested. Our experimental results provide no clear indication whether minimal narrative or rich narrative have a greater effect on learning. More research is needed to tease out whether narrative has an effect on learning. 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 our participants as they experienced the game play itself.

Nevertheless, our experiment suggests that educational games hold promise at effectively teaching people about cognitive biases, a fairly pernicious aspect of human information processing and decision-making. More research is needed to examine whether other game features might have a greater impact on learning.

ENDNOTES

[1] The FAE scale contained 10 items and had alphas of .70 (pretest), .89 (posttest), and .88 (retention). The CB scale contained 5 items and had alphas of .82 (pretest), .77 (posttest), and .75 (retention). The BBS scale contained 7 items and had alphas of .88 (pretest), .85 (posttest), and .85 (retention).

ACKNOWLEDGMENTS

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.

BIBLIOGRAPHY

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.

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

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 copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, *15*(4), 359-372.

Conle, C. (2003). An anatomy of narrative curricula. *Educational Researcher*, *32*(3), 3-15.

Dale, R., Kehoe, C. E., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15-28.

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

Dettori, G., & Paiva, A. (2009). Narrative learning in technology-enhanced environments: an introduction to narrative learning environments. In N. Balacheff et al. (Eds.) *Technology-enhanced learning, Part 1*, (pp. 55-69). New York: 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 & Research Development*, *54*(3), 245-263.

Eisner, E. W. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice.* Merril, N.J.: Prentice Hall.

Fischer, P., & Greitemeyer, T. (2008). A new look at selective-exposure effects: An integrative model. Current Directions in Psychological Science, 19, 384-389.

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

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Palgrave/St. Martin's.

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

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory, 14*(4), 311-327.

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 & Behavior*, *34*(5), 777-792.

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

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

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

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333-369.

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

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. Retrieved from http://www.cs.cmu.edu/afs/cs/user/michaelm/www/nid.

Nowak, K. L., Hamilton, M. A., & Hammond, C. C. (2008). The effect of viewer perceptions of avatar anthropomorphism and realism on judgments of credibility and homophily, and avatar choice. *PRESENCE* 2008 Proceedings from *11th Annual International Workshop on Presence Padova, 16-18 October.*

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.

Rahford, N. (2009). Overcoming cognitive bias in group decision-making. Retrieved from http://www.noahraford.com/files/raford_overcomingbias.pdf

Rieber, L. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Re-search and Development*, *44*(2), 43-58.

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

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 & Cognition, 30*(8), 1288-1296.

Shaw, A. (2010). Identity, identification and media representation in video game play: An audience reception study. (Unpublished Dissertation). University of Pennsylvania, Philadelphia.

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*(1), 221-225.

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

Jenkins, H. (2004). “Game design as narrative architecture.” First person: new media as story, performance, and game. Ed. Noah Wardrip-Fruin & Pat Harrigan. Cambridge, Ma.: The MIT Press.

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.

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: Routledge.