Article

Games and Culture
1-17

DOI: 10.1177/1555412016673262
gac.sagepub.com

The Design and Development of Serious Games Using Iterative Evaluation

Carl Symborski[1], Meg Barton[1], Mary Magee Quinn[1], James H. Korris[2], Karim S. Kassam[3] and Carey K. Morewedge[4]

Abstract

In this article, we report on a serious game development approach, characterized by combining theory-based design with an iterative development strategy guided by experimental test and evaluation. We describe two serious games that teach the mitigation of cognitive biases (human tendencies to commit systematic errors in thinking that lead to irrational judgments). Cognitive biases tend to be deeply ingrained and early attempts to reduce biases with training have met with little success. We address this training challenge using bias mitigation theory derived from the literature and an instructional framework to establish the educational content of each game. The mitigation effects of the games were measured through multiple experiment cycles, and multiple play-testing campaigns were conducted to inform instructional model and game design revisions. The final game versions achieved a medium-to-large training effect following a single play session.

Keywords

serious games, game design, iterative testing, training, cognitive bias

[1] Leidos, Arlington, VA, USA
[2] Creative Technologies, Inc., Los Angeles, CA, USA
[3] Carnegie Mellon University, Pittsburgh, PA, USA
[4] Boston University, Boston, MA, USA

Corresponding Author:
Meg Barton, Leidos, 4001 N Fairfax Dr., Arlington, VA 22203, USA.
Email: marguerite.r.barton@leidos.com

As human beings, we are faced with thousands of decisions every day, from the relatively inconsequential (What should I wear today?) to the momentous (Should I go back to school to get an advanced degree?). In practice, we do not have time to ponder every decision individually—to slowly weigh every pro and con of each one. Consequently, we rely on a number of heuristic decision rules that act as mental shortcuts, enabling us to make satisfactory decisions quickly and efficiently. While heuristics are generally effective, they have the potential to go awry, leading to illogical thought processes that produce irrational judgments and decisions. These systematic errors in thinking are known as "cognitive biases."

The Intelligence Advanced Research Projects Activity (IARPA) Sirius program was designed to investigate whether video games could be an effective tool for training players to recognize and mitigate their cognitive biases (IARPA Incisive Analysis Office, 2011). The literature on cognitive biases indicates that they are ubiquitous, deeply ingrained in the human psyche, and resistant to most forms of debiasing training (Morewedge et al., 2015). Given this training challenge, the IARPA Sirius program adopted the novel approach of applying "serious games" (i.e., games designed for purposes beyond pure entertainment, often to train or to educate the player) to the problem of cognitive biases. Serious games have been applied to numerous problem domains in recent years, from foreign language and culture skills acquisition (Johnson, 2007; Johnson, Vilhjalmsson, & Marsella, 2005), to promoting treatment adherence for cancer patients (Beale, Kato, Marin-Bowling, Guthrie, & Cole, 2007), to training conflict resolution skills (Grappiolo, Cheong, Togelius, Khaled, & Yannakakis, 2011).

The IARPA Sirius program targeted six specific cognitive biases relevant to intelligence analysis over two phases of the program. During the first phase of the Sirius program (i.e., Phase 1), we addressed three of the six biases: confirmation bias, the tendency to seek out and interpret evidence in a way that confirms rather than disconfirms a favored hypothesis (Nickerson, 1998); the fundamental attribution error, the tendency to make dispositional rather than situational attributions for a person's behavior (Gilbert, 1998; Jones & Harris, 1967); and bias blind spot, the tendency to recognize bias in others but not in ourselves (Scopelliti et al., 2015). In the second phase of the program (i.e., Phase 2), we addressed the remaining three biases: anchoring bias, the tendency to overweight a single piece of information when making a decision (Tversky & Kahneman, 1974); the representativeness heuristic, the tendency to judge probability based on appearances or what "seems right" (Kahneman & Tversky, 1972); and projection bias, the tendency to assume that others share our emotions, thoughts, and values (Epley, Morewedge, & Keysar, 2004; Robbins & Krueger, 2005).

The structure of the IARPA Sirius program was a key influence on the game development process. The program was divided into two phases, each of which focused on three biases, as described above. Within each phase, there were three experimental test cycles. During the program, multiple teams independently developed serious games to promote bias recognition and mitigation. Progress toward the program objectives was measured at each cycle of experimentation, and the best performing games were permitted to continue to subsequent cycles. We, Team Leidos,[1] were one of the two teams that completed the entirety of the Sirius program. This program structure, which allowed for iterative testing of the game efficacy across experiment cycles, influenced the initial game design as well as the game refinement approach that we adopted, which became a key factor to developing a successful game.

In this article, we describe our game design approach for two serious games to combat cognitive biases. The first game, developed during Phase 1 of the program, is titled *Missing: The Pursuit of Terry Hughes*; the second game, developed during Phase 2, is titled *Missing: The Final Secret*. With the knowledge that cognitive biases have traditionally been difficult to mitigate, we incorporated a theory-based cognitive bias framework (Barton et al., 2015; Symborski et al., 2014) and an instructional framework from the outset to optimize the training potential of the games, which we outline below. We then describe the iterative process by which we refined the games across experimentation cycles, guided by learning outcome and play-testing data.

Game Design Approach

As the two *Missing* games were envisioned to be companion training tools (*Missing: The Final Secret* acts as a sequel to *Missing: The Pursuit of Terry Hughes*), they were implemented in a similar style and genre. Both games are driven by a strong narrative and take on the style of an adventure game, with cognitive bias training activities woven into the gameplay. In the adventure game genre, the player is immersed in an interactive story, wherein she can explore and solve puzzles. We selected this genre because it would allow us to create a realistic, modern day game world; both *Missing* games take place in present day New York City. By presenting the game content in a realistic context, we sought to educate the player about biases in a way that would facilitate generalization of learning to the "real world." Figure 1 presents screenshots from the *Missing* games to illustrate the games' general aesthetic.

For each phase of the program, a cognitive bias framework linking the target biases for that phase, their causes, and mitigation strategies for those causes was developed based on relevant literature and current theory in the field (Barton et al., 2015; Symborski et al., 2014). The Phase 1 and Phase 2 cognitive bias frameworks were developed at the outset of each phase and were not altered once game development began. In contrast, our initial instructional framework was expanded and refined throughout the program. The cognitive bias framework and instructional model were foundational aspects of our game design and are described in greater detail in the following sections.

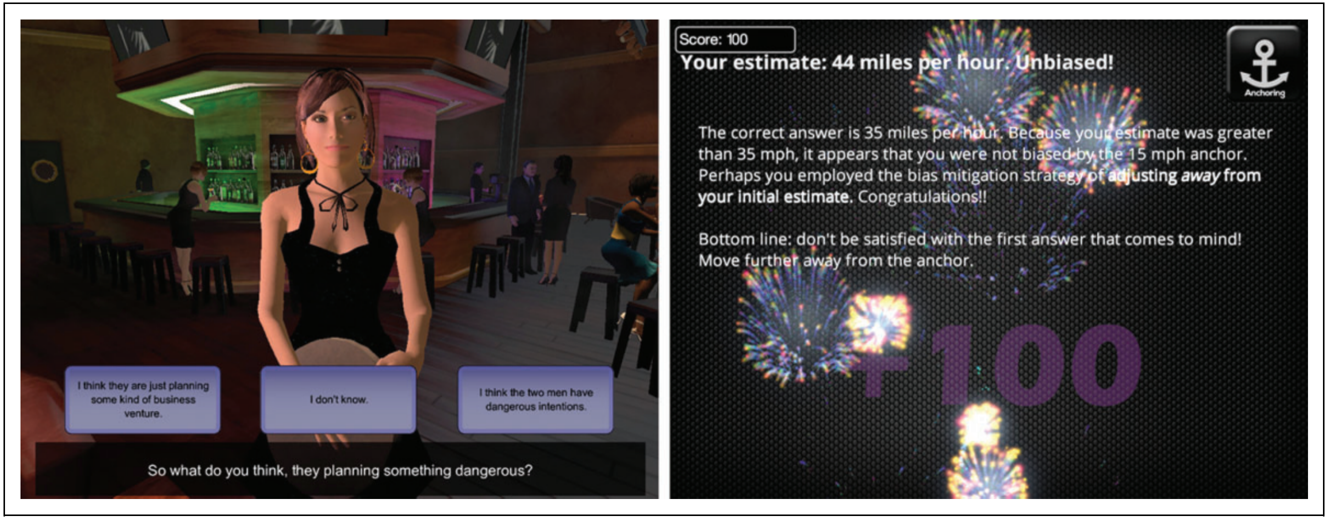

Figure 1. The player interacts with a game character during the third episode of *Missing: The Pursuit of Terry Hughes* (left); the player receives feedback on his or her in-game performance during the first After Action Review segment of *Missing: The Final Secret* (right).

Theory-Driven Approach: Cognitive Bias Framework

Historically, training participants to mitigate cognitive biases has proven exceptionally difficult (Croskerry, Singhal, & Mamede, 2013; Kahneman, 2011). To address this training challenge and optimize the teaching potential of our games, we developed a "cognitive bias framework" as the foundation for our game design. The cognitive bias framework underpinning both *Missing* games was based on current research on the underlying causes and mitigations for cognitive biases; specifically, the theory of dual-process systems of reasoning (Evans, 2007; Morewedge & Kahneman, 2010). This theory holds that there are two systems of reasoning that we employ when rendering judgments. The first, System 1 reasoning, is described as fast, automatic, intuitive thinking; the second, System 2 reasoning, is characterized as slower, more logical, and more deliberate thinking (Evans, 2007; Morewedge & Kahneman, 2010). The theory of dual-process systems of reasoning asserts that cognitive biases arise when the logical, rule-governed processes of System 2 reasoning fail to moderate the automatic and reactive processes of System 1 reasoning, resulting in irrational judgments (Morewedge & Kahneman, 2010).

Based on the principles of this theory, we set about creating cognitive bias frameworks that would link the target biases, their underlying causes (i.e., the System 1 reasoning processes that give rise to biases when left unchecked), and relevant mitigation strategies targeting those underlying causes (i.e., the System 2 reasoning processes that can correct the errors in thinking generated by overhasty System 1 reasoning processes). Separate frameworks were developed for Phases 1 and 2 of the program, specific to the target biases for each phase. For a detailed description of the Phase 1 and Phase 2 cognitive bias frameworks, please refer to Symborski et al. (2014) and Barton et al. (2015), respectively.

Theory-Driven Approach: Instructional Framework

In addition to the cognitive bias framework, we implemented an instructional framework to maximize the teaching potential of the games. The basic instructional framework, described below, was bolstered by a number of learning theory principles drawn primarily from the theories of andragogy (i.e., adult learning theory; Knowles, 1990) and constructivism. For further detail on how the principles of andragogy and constructivism were incorporated into the game design, please refer to Symborski et al. (2014).

Our instructional framework, which took the form of a Play–Teach loop, was a fundamental influence on the game structure that was adopted for both *Missing* games. The games are divided into three episodes, each of which contains a gameplay segment that takes place in a 3-D virtual environment and an After Action Review (AAR) segment. During the gameplay segment, the player is immersed in the narrative and encounters "bias vignettes," which elicit the target biases and provide opportunities for the player to commit or avoid those biases in a realistic environment (Play). Following the gameplay segment of each episode, there is an AAR in which the player receives instruction on the biases, is taught the appropriate mitigation strategies, and is given feedback regarding his or her performance on the bias vignettes (Teach). The Play–Teach cycle begins anew as the player enters the second and third episodes of the game, which afford the opportunity for the player to apply what he has learned in the previous episodes and learn additional content taught in these episodes.

This Play–Teach loop was synergistic with the cognitive bias framework. The cognitive bias framework dictated that players should be exposed to common causes of the biases, such that the corresponding mitigation strategies for those causes could be presented to players during the game. Accordingly, players experience cognitive bias elicitation paradigms (i.e., causes) during the gameplay segment of each episode (Play) and are taught the appropriate mitigation strategies during the AARs (Teach). They can then better recognize when biases are being elicited and implement the mitigation strategies during the subsequent episodes.

The decision to begin the game with a “Play” segment, rather than initiating with a “Teach” segment, was a deliberate design choice that was motivated by two primary considerations. First, since players entering the game world are potentially naive to the teaching purpose of the game, they are more likely to exhibit biases during the initial Play segment; this renders the teaching content that follows more personally relevant to the player. Second, by hooking the player’s attention at the outset of the game, we hoped to promote greater engagement with the storyline and commitment to learning.

While the cognitive bias framework remained constant for each phase of the program, the Play–Teach instructional framework evolved across experiment cycles during Phase 1 as well as between Phase 1 and Phase 2. This evolution process, along with the cross-cycle game refinements that were made throughout the Sirius program to produce the final versions of the *Missing* games, is described below.

Game Refinement Through Testing

Team Leidos was one of the several teams independently developing serious games for the mitigation of cognitive biases as part of the IARPA Sirius program. IARPA defined performance objectives by which the effectiveness of the developed games to teach and mitigate cognitive biases would be measured. As described previously, the IARPA Sirius program structure consisted of two main program phases, each of which was composed of a series of three experimental test cycles. All teams performed their own test protocols, reporting results at the end of each test cycle. Further, a government-led verification and validation team performed their own independent evaluations to validate results put forth by each team. Following this structure, progress toward achieving the IARPA-specified performance metrics was measured at each cycle and the highest performing games continued to subsequent cycles.

This program structure significantly influenced the game development strategy that we selected. Given that our team could be dropped from the program at any cycle, we put emphasis on getting all of the educational content into the game in time for the first round of testing in each program phase. We further prioritized the implementation of our theory-based cognitive bias and instructional frameworks over fielding fully polished game graphics and mechanics. The intent of this strategy was to ensure that our game performance would be as close to the program goals as possible in the first test cycle of each phase.

After demonstrating that the game had made strong progress toward reaching program goals during the first test cycle of Phase 1 and Phase 2, we utilized the subsequent cycles to refine the games. The game refinement process was informed by two primary data sources: first, the efficacy results from each experimental test cycle and, second, the results of a play-testing campaign to gather feedback on the user experience. Using this strategy, refinements were made to the game with the objective of identifying the game content that would optimize learning. In the following sections, we review the methods by which we gathered the experimental test data and the play-testing data that were central to our game revision process.

Cycle-Testing Method

For each experiment cycle, we conducted a test campaign in which participants would arrive at our lab, complete a pretest, play the game,[2] then complete an immediate posttest. To evaluate longitudinal retention of cognitive bias knowledge and mitigation, participants were e-mailed a personalized link to a follow-up test 8 weeks (Phase 1) or 12 weeks (Phase 2) after the completion of the lab session. Participants were recruited either in the Pittsburgh, PA (Phase 1 and Phase 2), or Boston, MA (Phase 2), areas. Our recruitment goal for each cycle was 280 people, with most participants attracted from local colleges and universities.

For Phase 1, standardized measures of confirmation bias, the fundamental attribution error, and bias blind spot were developed by Carnegie Mellon University (Morewedge et al., 2015). For Phase 2, standardized measures of anchoring bias, the representativeness heuristic, and projection bias were developed jointly by a MITRE-led team consisting of researchers from Team Leidos, Applied Research Associates, and the University at Albany-State University at New York (Morewedge et al., 2015). Both the Phase 1 and Phase 2 assessment instruments consisted of two sections: one section covered the recognition and discrimination of the target biases (i.e., bias knowledge) and the other covered bias mitigation.

Our analysis of the data was guided by the IARPA Sirius program research question: are the games effective in teaching bias recognition and mitigation, and is the training effect retained over time? In order to address this question, we used *t*-tests to assess whether the differences from pretest to posttest and from pretest to follow-up were statistically significant. Further, we calculated percentage improvements and effect sizes to evaluate the magnitude of the games' bias mitigation capabilities. Results of this analysis were used to guide revisions to the game in preparation for subsequent test cycles.

Play-Testing Method

Our play-testing method was composed of two separate protocols: an informal play-testing protocol and a formal play-testing protocol. Informal play testing was conducted by members of Team Leidos and was primarily geared toward comprehensive testing of the game features to elicit any bugs that might require repair, akin to the "game breaking" method (Olsen, Procci, & Bowers, 2011). Conversely, the purpose of formal play testing was to gather feedback on the usability and playability of the games from volunteers outside of the project team. In conducting formal play testing, we hoped to gather data that could allow us to repair any game issues that might interfere with players' learning outcomes. Formal play testers were recruited from outside of the project team and were specifically directed *not* to deliberately seek out software bugs; instead, they were asked to immerse themselves in the games and report back on any issues with the game content, mechanics, or user experience. The formal play-testing protocol consisted of an introductory session, in which participants were informed of the objectives of the play-testing session; a gameplay session with affordances for data collection; a post-play test questionnaire with items assessing gameplay, mechanics, the game story, and usability; and a focus group debriefing session. Full details of our play-testing protocols are described in Quinn et al. (2013). Play-testing results were analyzed and reported to the game development team and used to guide revisions to the game in preparation for the subsequent test cycles.

Missing: The Pursuit of Terry Hughes—Phase 1, Experiment Cycles 1 Through 3

The Phase 1 game, *Missing: The Pursuit of Terry Hughes*, was developed from scratch, and we were not entirely certain at the outset what game genre and mechanics would perform well enough to meet the IARPA performance criteria. As discussed, we first focused on fully implementing our defined cognitive bias and teaching frameworks in the Cycle 1 game, then used data derived from play-testing and cycle-testing results to guide subsequent revisions. We found that the Cycle 1 game performed reasonably well, making good strides to meet the overall program objectives for mitigation of the fundamental attribution error (−55.2% at posttest[3] and −49.2% at follow-up) and bias blind spot (−44.9% at posttest and −28.9% at follow-up); however, the game performed poorly for confirmation bias (−2.2% at posttest and −0.9% at follow-up). With two additional build-test cycles remaining, we sought to determine what changes should be made to improve the game performance overall, with a particular focus on game revisions that would improve our performance for confirmation bias.

Revisions based on play-testing data. Based on play-testing feedback, changes were implemented to the game mechanics and visual presentation of information to improve the player experience. The goal of these revisions was to eliminate any game-induced barriers to learning. The primary game enhancements based on play-testing feedback fell into one of two areas: player orientation or usability improvements.

Player orientation. One theme that became apparent during the analysis of our play-testing data was that players found it important to orient to the game world before commencing with gameplay. Some play testers expressed concern that the game did not provide them with sufficient opportunity to fully orient to the game story before entering the game environment. The game begins with an opening video, intended to set the stage and draw the player into the story and, therefore, the game. However, we observed that some players did not anticipate that the opening video would contain any important content or were not ready to engage with the game as it began (e.g., because they were still fiddling with their headphones or the volume control or were adjusting their sitting position). Many play testers reported regretting that they had "missed something" in the opening video which might have been relevant to playing the game. To resolve this issue, we added an option to replay the opening video upon its completion. This allowed all players the chance to orient to the game story before beginning to play the game.

In a similar vein, some play testers reported that they did not know who they were, why they were there, or what they were supposed to do upon entering the game world. To address this issue, we added a text-based "who am I" introduction screen, which appears after the opening video, to clarify these points and provide player orientation.

Usability improvements. Players who become annoyed at perceived flaws in a game user interface have the potential to become less focused on learning and more focused on the game software implementation. This can be a particular challenge for serious games, which are generally implemented with modest development budgets compared to popular entertainment games. We had the benefit of multiple rounds of play testing to gather specific usability improvement suggestions, which we then implemented in order of priority (i.e., we focused first on the usability improvements that were most frequently requested by play testers). One key improvement was to implement a "Next" button during the AARs, which advanced players through the AAR screens as they clicked the button. The Next button gave the player agency to self-pace through the game's teaching and feedback material.

Some play testers struggled with navigating through the game's virtual world environment. Not all of our participants were experienced gamers. Indeed, a game for training should not require gaming expertise, if the game is to be applicable to the widest possible audience. To improve the user experience, we added a tutorial on how to move about the game environment. Navigation was further improved by tuning the speed and sensitivity of the controls and adding a "go to" option on the clickable objects in the environment. Additionally, in the first episode of the game where the player navigates around an apartment with multiple rooms, we added a map pop-up that allowed the player to click on a room in the apartment on the map and be navigated there automatically.

In some cases, removing or simplifying user options can improve usability. Our game includes a "smartphone" with which players communicate via text messages, take pictures, and record their observations. After observing play testers struggling to learn how to use the smartphone, we simplified its appearance when displaying text messages and made taking and sending pictures more intuitive.

Revisions based on cycle-testing data. We relied on the efficacy testing results of Cycle 1 and Cycle 2 to guide changes to the bias content and instructional model in the game. Initial efficacy testing showed that our game was performing well for the fundamental attribution error and bias blind spot, but poorly for confirmation bias. To address this, we added additional content to the game and expanded our instructional model for this bias.

In order to increase players' exposure to confirmation bias and thereby increase opportunities for learning, we added four bias vignettes focused on improving the ability of users to identify and mitigate confirmation bias. These supplementary bias vignettes gave players further exposure to potential causes of confirmation bias during the gameplay segments of each episode, paired with additional feedback and teaching content on confirmation bias mitigation strategies during the AARs.

Additionally, we expanded the instructional framework (Play–Teach) for confirmation bias to include a "Test" paradigm. The Test paradigm was inserted into the confirmation bias segments of the AARs to provide additional active instruction on the bias mitigation strategies. After teaching confirmation bias definitions and providing feedback to players on their game performance, players were given additional confirmation bias challenges. These practice problems provide an opportunity for players to apply their knowledge and receive additional feedback before returning to the next episode of the game. With this revision, we had implemented a Play–Teach–*Test* loop within each episode for confirmation bias.

Our Cycle 3 results demonstrated the improvement these changes had in game efficacy performance. The game significantly improved participants' ability to recognize and discriminate between the biases, both at immediate posttest, $t(159) = 13.33, p < .001$, and at follow-up, $t(129) = 5.87, p < .001$. Overall, game training effectively reduced cognitive bias immediately and 2 months later, $t(159) = 21.24$, $p < .001$, and $t(129) = 12.65$, $p < .001$, respectively. Debiasing effect sizes (Rosenthal & Rosnow, 1991) for overall bias were large for the game, both immediately (-46.3%; $d_{\text{pre–post}} = 1.68$) and 2 months following the training session (-34.8%; $d_{\text{pre–follow-up}} = 1.11$). Debiasing effect sizes were also large for each of the three target biases at immediate posttest (confirmation bias: −28.5%, $d_{\text{pre–post}} = 1.09^{***4}$; fundamental attribution error: −58.9%, $d_{\text{pre–post}} = 1.12^{***}$; bias blind spot: −57.7%, $d_{\text{pre–post}} = 0.98^{***}$) and large to medium at follow-up (confirmation bias: −15.6%, $d_{\text{pre–follow-up}} = 0.58^{***}$; fundamental attribution error: −40.8%, $d_{\text{pre–follow-up}} = 0.72^{***}$; bias blind spot: −61.1%, $d_{\text{pre–follow-up}} = 0.89^{***}$).

Missing: The Final Secret—Phase 2, Experiment Cycles 4 Through 6

The Phase 2 game, *Missing: The Final Secret*, continues the adventures began in the Phase 1 game while providing instruction on the three cognitive biases of interest for Phase 2 of the program: anchoring bias, the representativeness heuristic, and projection bias. In developing the Phase 2 game, we carried forward lessons learned from the Phase 1 research. Specifically, we implemented the revised Play–Teach–Test instructional model, shown to be effective for confirmation bias in Phase 1. This was implemented for all three biases in the Cycle 4 version of the game (i.e., the first Phase 2 game version used for experimental testing). The Cycle 4 game version also included a simple student model, whereby the game offers additional practice questions in the Test portion of the instructional model for those players who struggle with the lessons.

Additionally, we carried forward the player orientation and usability enhancements developed during Phase 1 testing. For the Phase 2 game, we ensured that players are given an explanation of their own backstory, the ongoing events within the game world as they enter it, and why they are involved with those events. This was implemented as a player orientation screen that appears immediately following the opening video. Further, player navigation controls were calibrated similarly to those used in the Phase 1 game and a navigation tutorial was implemented as in the Phase 1 game. Finally, we carried forward the Next button in the AARs to give the player agency to self-pace through the teaching material.

Following the successful Phase 1 strategy, we first focused on implementing our cognitive bias and instructional frameworks in the Cycle 4 game, then used the results of play-testing and cycle-testing activities to guide subsequent game revisions. The Cycle 4 game performed well, particularly for projection bias (−41.1% at posttest and −24.7% at follow-up) and the representativeness heuristic (−67.7% at posttest and −53.9% at follow-up). Whereas in Phase 1, the initial game version struggled to mitigate confirmation bias, in Phase 2, the mitigation of anchoring bias proved to be the most challenging (−15.3% at posttest and −15.0% at follow-up). Once again, we used the two remaining build-test cycles to make data-based revisions to the game that we hoped would maximally improve game performance overall. In particular, we focused on revisions that could improve the game's ability to mitigate anchoring bias.

Revisions based on play-testing data. Starting with play-testing feedback, changes were implemented to game mechanics, assets, and content to eliminate any barriers to learning. Game enhancements based on play-testing feedback fell into one of three areas: fixing art asset bugs, increasing player agency, or revising player feedback.

Fixing art asset bugs. Resolving software problems is an expected and necessary part of any software development effort. Based on play-tester comments, we determined that some graphical bugs in the game were distracting players from being fully engaged, highlighting the importance of smooth-running, glitch-free software in supporting learning from serious games. A specific example of a graphical bug that play testers found distracting was an instance in which a game character's hair would change in graphical resolution during one of the bias vignettes, creating the appearance that parts of her hair appeared and disappeared as she spoke. Also distracting were occasional cases where game characters' clothing became distorted during animation sequences, causing underlying skin to show through. We invested time and effort in repairing these types of bugs to reduce potential distractions that could dampen player learning.

Increasing player agency. Player navigation in the Cycle 4 game version was pre-programmed and tightly bound to the story narrative. Players were essentially "on rails" as they played the game, with limited ability to move about on their own. This was a departure from the Phase 1 game, where players had opportunities to move about the playfield freely, and play-testing feedback regarding the lack of player agency in the Cycle 4 game was universally negative. As this was considered a potential detractor from player engagement, we expanded the game in the Cycle 5 and Cycle 6 versions to allow players to freely explore the playfield as part of completing tasks. Further, we added additional clickable objects for players to discover while exploring.

Revising player feedback. Our instructional framework includes an AAR, wherein players are given feedback on their in-game performance. During play testing, we noted that many testers objected to being given feedback that they had made a biased response and would thereafter be less receptive to any feedback from the game. To address this, we revised the feedback text to soften the news that a player had given a biased response. Specifically, instead of telling players that they were biased, the feedback text would focus on the *answer* (vs. the person) *possibly* having been impacted by bias (vs. the answer definitively having been impacted by bias). Similarly, feedback in our game often depends on how close a player is to the correct answer, particularly with anchoring bias questions. Feedback was revised to expand the window of what is considered "close" to the correct answer and to provide positive feedback when the player is close to the correct answer. The intent of this revision was to encourage players to stay focused on learning and to reduce frustration due to the perception of overly critical feedback.

Revisions based on cycle-testing data. Following the same experimental approach used in Phase 1, for Phase 2, we relied on efficacy testing results of the first two test cycles (Cycle 4 and Cycle 5) to guide changes to the cognitive bias content in the game. As mentioned, initial efficacy testing showed poor mitigation performance of our game for anchoring bias. To address this, we incorporated additional anchoring bias content into the game and replaced one of our anchoring bias practice examples with an improved example. We also added one additional projection bias practice example to help improve efficacy results for that bias.

Concerned about the overall balance of playing to teaching in the game, given the additional anchoring and projection bias content, we made several changes to reduce the duration of the AARs. We first optimized the control logic to present fewer examples to players performing well, which constituted a revision to our student model. We also streamlined player feedback text, improving readability by trimming redundant and excess wording. Finally, we enabled the Next button for each AAR screen to be shown after several seconds rather than requiring the player to listen to the narrator read all of the on-screen text before continuing.

Our Cycle 6 results demonstrated the improvement these changes had in the efficacy performance of the game. The game produced significant improvements in recognition and discrimination of the three biases at immediate posttest, $t(187) = 17.59$, $p < .001$, an effect that was maintained at the 3-month follow-up, $t(142) = 7.14$, $p < .001$. Overall, game training effectively reduced cognitive bias immediately, $t(187) = 28.4$, $p < .001$, and 3 months later, $t(142) = 17.8$, $p < .001$. Debiasing effect sizes for overall bias were large for the game at immediate posttest (-40.8%; $d_{pre\text{–}post} = 2.08$) and at follow-up ($-29.2\%$; $d_{pre\text{–}follow\text{-}up} = 1.49$). Additionally, effect sizes were large for each of the three target biases at immediate posttest (anchoring bias: -17.8%, $d_{\text{pre–post}} = 1.10$***; representativeness heuristic: -61.7%, $d_{\text{pre–post}} = 1.72$***; projection bias: -43.0%, $d_{\text{pre–post}} = 1.46$***); these effects remained large at follow-up (anchoring bias: -16.0%, $d_{\text{pre–follow-up}} = 0.90$***; representativeness heuristic: -46.3%, $d_{\text{pre–follow-up}} = 1.23$***; projection bias: -25.4%, $d_{\text{pre–follow-up}} = 0.81$***).

Conclusions

The IARPA Sirius program was an ambitious and innovative research effort to determine whether serious games could be used to train the recognition and mitigation of cognitive biases. To tackle this training challenge, Team Leidos developed theory-based cognitive bias and instructional frameworks as the foundation of our games' content and structure. We then refined the games based on play-testing and efficacy-testing data over each phase of the program. This iterative game design approach allowed us to optimize our game content, usability, and teaching strategies over time to produce the most effective final game versions, within the time and budgetary constraints of the program.

Many of the game revisions that we implemented, guided by play-tester feedback and the results of our game efficacy testing, suggest useful design principles for those seeking to develop effective serious games. Our play-testing data indicate that providing players with an adequate orientation to the game world before commencing with gameplay is important, as are simple, intuitive navigation mechanics and game controls. Small glitches in game graphics may seem inconsequential but can prove disruptive to player focus. Presenting feedback in a manner that is not overly harsh or critical may prevent alienating players and encourage them to remain receptive to the game's teaching. Allowing players to self-pace through teaching material may be beneficial for learning outcomes as well. Additionally, maximizing opportunities for player agency in the game promotes engagement.

Our efficacy testing data informed the game revision process by highlighting areas where we could improve the teaching content for particular topics. Efficacy testing results prompted us to incorporate additional content and refine the teaching strategies for confirmation bias in Phase 1 and anchoring bias in Phase 2. In both cases, players' bias mitigation capabilities increased following the expansion and refinement of the teaching content.

At the conclusion of the Sirius program, our results indicate that it is indeed possible to train players to recognize and mitigate their cognitive biases through a serious game intervention. The structure of the program, which allowed for an iterative design approach informed by empirical data, was essential to its success. Other game developers may benefit from this type of approach when developing their own serious games.

Authors' Note

The views and conclusions contained in the article 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.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory (AFRL) contract number FA8650-11-C-7175.

Notes

1. Team Leidos consisted of researchers from Leidos, Boston University, Carnegie Mellon University, and Creative Technologies Incorporated.

2. During each test cycle, some participants were randomly assigned to play the game, while others were randomly assigned to watch an educational control video. This aspect of our study design, which enabled us to compare the efficacy of the game to a more traditional teaching method (i.e., the control video), is described in greater detail in Symborski et al. (2014), Barton et al. (2015), and Morewedge et al. (2015).
3. That is, a 55.2% reduction in bias at posttest from the pretest score.
4. ***Significance at $p < .001$.

References

Barton, M., Symborski, C., Quinn, M., Morewedge, C. K., Kassam, K. S., & Korris, J. H. (2015). The use of theory in designing a serious game for the reduction of cognitive biases. *Proceedings of DiGRA 2015 Conference: Diversity of Play*. Retrieved from http://www.digra.org/wp-content/uploads/digital-library/164_Barton_etal_The-Use-of-Theory-in-Designing-a-Serious-Game.pdf

Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., & Cole, S. W. (2007). Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *Journal of Adolescent Health*, *41*, 263–270.

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 2: Impediments to and strategies for change. *BMJ Quality and Safety*, *22*, ii65–ii72. doi:0.1136/bmjqs-2012-001713

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*, 760–768.

Evans, J. S. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment* (*Vol. 3*). New York, NY: Psychology Press.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindsey (Eds.), *The handbook of social psychology* (Vol. *2*, pp. 89–150). Hoboken, NJ: John Wiley.

Grappiolo, C., Cheong, Y. G., Togelius, J., Khaled, R., & Yannakakis, G. N. (2011, May). Towards player adaptivity in a serious game for conflict resolution. In *Games and Virtual Worlds for Serious Applications (VS-GAMES 2011), Third International Conference* (pp. 192–198). doi:10.1109/VS-GAMES.2011.39

Intelligence Advanced Research Projects Activity Incisive Analysis Office. (2011, April 25). *IARPA Broad Agency Announcement: Sirius* (IARPA-BAA-11-03). Washington, DC: Federal Business Opportunities.

Johnson, W. L. (2007). Serious use of a serious game for language learning. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 67–74). Fairfax, VA: IOS Press.

Johnson, W. L., Vilhjálmsson, H. H., & Marsella, S (2005, May). Serious games for language learning: How much game, how much AI? In C. K. Looi, G. McCalla, & B. Bredeweg (Eds.), *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology* (Vol. 125, pp. 306–313). Amsterdam, Netherlands: IOS Press.

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

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

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Knowles, M. S. (1990). *The adult learner: A neglected species* (4th ed.). Houston, TX: Gulf Publishing Company.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*, 435–440.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*, 129–140. doi:10.1177/2372732215600886

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

Olsen, T., Procci, K., & Bowers, C. (2011). Serious games usability testing: How to ensure proper usability, playability, and effectiveness. In A. Marcus (Ed.), *Design, user experience, and usability: Theory, methods, tools and practice* (pp. 625–634). doi:10.1007/978-3-642-21708-1_70

Quinn, M. M., Symborski, C., Barton, M., Korris, J., Falstad, T., & Granato, S. (2013). Methodology for playtesting serious games: A case study using a mixed method approach. *Games Innovation Conference (IGIC), 2013 IEEE International* (pp. 222–227). doi:10.1109/IGIC.2013.6659130

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*, 32–47.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill.

Scopelliti, I., Morewedge, C. K., McCormick, E., Min, H. L., Lebrecht, S., & Kassam, K. S. (2015). Bias blind spot: Structure, measurement, and consequences. *Management Science*. doi:10.1287/mnsc.2014.2096

Symborski, C., Barton, M., Quinn, M., Morewedge, C., Kassam, K., & Korris, J. H. (2014). Missing: A serious game for the mitigation of cognitive biases. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Retrieved from http://www.iitsec.org/about/publicationsproceedings/documents/bp_trng_14295_paper.pdf

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

Author Biographies

Carl Symborski is a chief engineer at Leidos where he leads diverse teams of scientists, engineers, and subcontractors to design and deliver innovative solutions for DARPA, IARPA,

and others. Recent programs include research in insider threat detection, developing serious games, and researching social dynamics in virtual worlds.

Meg Barton is a life scientist at Leidos with a background in human subjects research and usability testing. She has worked on gaming-related research projects at Leidos for the last 5 years.

Mary Magee Quinn is a senior behavioral scientist at Leidos and a nationally recognized expert in applied behavior analysis and functional behavioral assessment. She has more than 25 years of experience in the clinical analysis of behavior across a variety of domains and populations and has extensive experience in the field of education.

James H. Korris is CEO and president of Creative Technologies Incorporated (CTI). Named as one of Military Training Technology's 2011, 2012, and 2013 Top 100 Simulation and Training companies, CTI is at the forefront of immersive virtual simulation development.

Karim S. Kassam is an assistant professor in the Department of Social and Decision Sciences at Carnegie Mellon University. His research focuses on emotion and decision-making.

Carey K. Morewedge is an associate professor of Marketing at the Questrom School of Business, Boston University, where his research includes the cognitive and affective processes that are involved in judgment and decision-making.