Heuristica: Designing a Serious Game for Improving Decision Making

Gwen Mullinix, Oliver Gray, Juan Colado, Elizabeth Veinott, James Leonard, Elizabeth Lerner Papautsky, Chris Argenta, Marcia Clover, and Steven Sickles
ARA, Inc.
Dayton, OH, U.S.

Chris Hale and Elizabeth Whitaker
Georgia Technology Research Institute
Atlanta, GA, U.S.

Edward Castronova, Peter M. Todd, Travis Ross, Jared Lorince, Jared Hoteling, and Sharon Mayell
Indiana University
Bloomington, IN, U.S.

Robert R.Hoffman
Institute for Human and Machine Cognition
Pensacola, FL, U.S.

Olivia Fox and John Flach
Wright State University
Dayton, OH, U.S.

***Abstract*— This paper describes the design process and development of a 3D immersive serious game, Heuristica. The objective of this video game is to train players to improve their decision making by mitigating cognitive biases in an engaging and effective way. Heuristica is the result of three development and empirical evaluation cycles over 18 months. Several game features have been tested, and only those that improved learning while maintaining engagement have been retained in the latest version of the video game. These include reward, real time feedback, and game customization. Our development and playtesting process is summarized, and the implications for designing training are described.**

Keywords—video games; serious games; simulation; bias; training; cognition; experiment; decision making

I. Introduction

Video games, in particular serious games, can improve learning and training across a variety of domains, including medicine, humanitarian, military, science, and education. Not only are video games engaging, but their potential for education is far reaching and under realized [1, 2]. In this paper, we describe the 18-month development of a 3D immersive serious game, called Heuristica, which was designed to improve specific aspects of players' decision making skills. Heuristica was built with the Unreal Engine 3, an award winning game engine commonly used by commercial entertainment game companies.

Heuristica was built as a multi-staged modular puzzle game. It is multi-staged because three distinct game phases (tutorial, training, and action) provide support for players of all levels of video game experience. The tutorial provides an introduction to the narrative, control scheme, and rules; the training phase continues training and gauges the players' skill level; and the action phase tests the players' cognitive abilities. Heuristica is modular because in the latter two phases of the game, the player completes the game in through a series of modules, called Learning Opportunities (LOs). These modules are mini games that engage the player in specific decision-making and problem-solving activities and the modular system ties all of the encounters together through a non-linear narrative that can be experienced by the user in different orders, yet still relates each encounter back to the overall story.

Heuristica is also a platform for experimental research. We used psychological theory and game design principles as a basis for our design decisions and use empirical feedback on learning and engagement to iteratively improve the video game. As an experimental platform, Heuristica has been used to empirically evaluate the effects of different game features (such as 3rd person perspective, reward, real time feedback, and customization) on learning outcomes in three different experiments with more than 600 participants. The results from our most recent experiment, presented in another paper at this conference [3], demonstrate that the game improves players' ability to mitigate bias. Overall, participants improved their knowledge of cognitive biases and increased their ability to mitigate them after playing Heuristica and these learning effects were still present after 8 weeks. Improved knowledge and increased ability to mitigate cognitive bias were measured by the difference between the pretest and a posttest scores (using different equivalent forms of a 41 item cognitive bias test). Participants improved their knowledge of the cognitive bias by 23% after playing Heuristica, compared to 16% for a more traditional training approach (i.e., an

---

For more information, please contact us at www.virtualheroes.com or the principal investigator, Beth Veinott, bveinott@ara.com.

engaging instructional decision video). After an 8week delay, participants retained their learning. Not only did players increase their knowledge of the cognitive biases, but they also were able to mitigate them. Participants reduced their cognitive bias, on average by 14% immediately after playing Heuristica, and performed even better after an 8-week delay (during which the game was not played). In contrast, participants who watched the instructional video reduced their cognitive bias by less than 5%%, but did not retain that learning. After an 8-week period of not playing Heuristica, we found that our game participants retained their ability to mitigate confirmation bias by 10% on average, fundamental attribution error by 21%, and bias blind spot by 50% over where they had started. More detail about these results from our most recent experiment is presented in another paper at this conference [3]. This has allowed us to incorporate experimentally validated game variables in subsequent versions of Heuristica.

In this paper, we describe the development of Heuristica. Specifically, we: 1) describe some of the theories underlying Heuristica's content and training approach, 2) review the iterative process used to develop Heuristica, 3) provide details of several game LOs and in game behavior, and 4) detail some latest game features.

A. Heuristica's training objectives

Heuristica was designed to teach players to recognize and mitigate three cognitive and social biases: confirmation bias, fundamental attribution error, and bias blind spot. *Confirmation bias* [5] refers to the tendency to favor confirming information over disconfirming information when searching for information or when testing hypotheses [6,7]. *Fundamental Attribution Error* occurs when individuals prefer personality explanations for others' actions, while neglecting the importance of the situation in determining another person's behavior [8]. For example, one might assume that Joe is late to a meeting due to personality traits (e.g., he is inconsiderate) without considering situational causes (e.g., traffic, or his bus arrived late). *Bias blind spot* is metacognitive tendency for a person to report that others are more susceptible to bias than oneself [10].

Our learning objectives included improving the knowledge, recognition, and mitigation of these biases. Video games have the power to teach through experiential learning and declarative learning by allowing players to fail, providing hints, and then allowing the player to try again [1]. Players can fail in a safe environment and revise their strategies to understand an experience or system. This allows players to not only learn from doing, but also increases the likelihood of practice. To increase engagement, and thus motivate the players to practice, we provide players with clear short and long-term goals, which tie together through the narrative of Heuristica. Even though the narrative is non-linear, the story of Heuristica unfolds as the players' progress through each module and stage. Play is structured as a single act with rising tension and a climax after players complete a number of modules determined by the intelligent tutor (student model). In building the narrative, we use theory from media psychology to provide players with goals that are embedded in the narrative and motivationally relevant to humans in general – e.g. survival, accumulation of resources, relationships [11]. Along with short-term goals that fuse into a long-term narrative, we also provide players with feedback to help them understand success and failure and we specifically provide two types of feedback to achieve this goal. The first provides information about the player's state in the game – it is information about the success or failure of a player's actions and allows a player to adjust her strategy through the course of play. The second type of feedback is focuses solely on positive reinforcement of goals and provides rewarding messages and experience points when players successfully accomplish a goal at various timescales. In addition to designing an engaging experience, we also designed Heuristica to allow players to practice mitigating cognitive biases in a wide range of situations. This approach generalizes the learning and reinforces it in long-term memory, thereby increasing the likelihood that a player would transfer what they learned in the game to a novel situation inside or outside of the video game.

II. Game description

In the Heuristica game, the player assumes the role of a human astronaut on a space station where a new starship crewed by a team of androids is about to launch the first Faster than Light expedition. A small team of humans must compete to win command over the androids on-board. The player is competing for one of these spots, and has made it to the final round of interviews. A core difference between the minds of androids and humans exists: where humans apply a set of cognitive and social heuristics, leaving them vulnerable to biases, androids use computation. To perform successfully alongside the androids, the player must demonstrate the ability to recognize and mitigate cognitive biases.

Heuristica's gameplay is driven by exploration, puzzle solving, and a narrative that integrates the different LOs in a space station context. The game is split into three stages: tutorial, training, and action. During the tutorial stage, the player learns basic gameplay controls and features. In the training stage, the game places each player in situations in which each of the cognitive biases might occur. After the player experiences a situation in which they may show cognitive bias, the game either teaches or reinforces a strategy for mitigating that bias (e.g., take a different perspective, consider a wider range of information). A non-player character called the Evaluator, a cyborg, acts as a guide or mentor during the training stage. At the end of this candidate training phase, a crisis occurs and the player is thrust into the action phase. In the action phase, the player's goal is to identify the cause of the crisis (e.g., enemy, mechanical) and either address it, or successfully escape the ship. The action phase is comprised of various situations that facilitate learning by presenting new problem solving opportunities. Players must then evaluate and analyze the information to make an accurate assessment to solve the problem. The player encounters new situations where bias could occur but under greater threat, and practices bias recognition and mitigation in this new situation. We describe a selection of Heuristica LOs below.

A. *Searching for and Weighting Information: Learning Opportunity for Confirmation Bias*

Confirmation bias can occur when people search for new information, or when they consider the available information. For this emergency medical-based LO (Fig. 1), players learn about this situation by developing a hypothesis about a patient's diagnosis based on the patient's symptoms (e.g., evidence). Specifically, non-player characters (NPCs, presented here as patients) have come in contact with one of two fictitious chemicals (Cyclodene and Morthium) during a hazardous chemical leak. The player must determine which chemical inflicted the patient's injury in order to treat them before they suffer serious effects. Players compare patient symptoms to chemical contact symptoms to properly diagnose the patient. Players are told that one chemical was found in greater amounts than the other (predisposing the players to assume one symptom set is more prevalent than the other). Each symptom the player chooses to query yields confirming, disconfirming or non-diagnostic information for a particular diagnosis. Consequently, players' successive choices can reflect a preference for confirming or disconfirming information. Players may only perform a limited number of queries. Therefore, to perform well on this task and advance in rank, players must learn which information to ignore, because it is non-diagnostic (same for both patients), and search for information that disconfirms the player's hypothesis. The player receives feedback in the form of patient responses, which encourages acknowledgement and understanding of how confirmation bias may influence the player's search strategies and decision making. Players exhibited the confirmation bias in this learning opportunity. In the first trial of this task, 64% of the players exhibited the confirmation bias and it went down to 33% over the course of the game. By the action phase, 20% of players exhibited confirmation bias. .

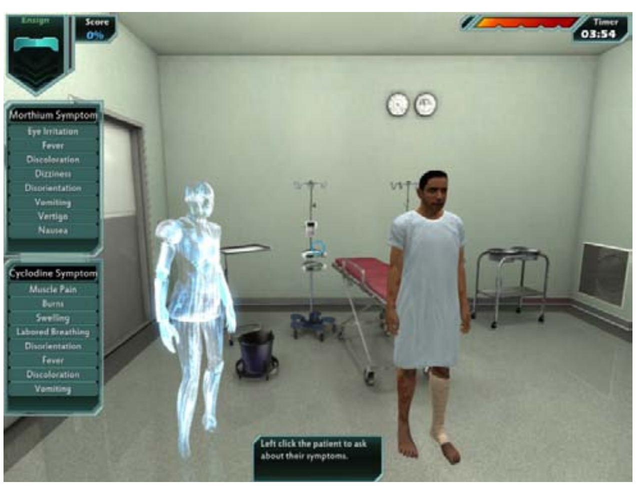

Fig. 1. Evidence Search and Weight in Diagnose

B. *Evaluating teamwork: Learning Opportunity for Teaching about the Fundamental Attribution Error*

In this LO (Fig. 2), which addresses fundamental attribution error, the player must work as part of a team (with an NPC) to successfully complete a task. The player must upload files to a server, using virtual agents called datadroids, to protect the information during a cyber attack. Datadroids are randomly attacked and destroyed by the enemy. After the player successfully uploads the files (or fails by crashing the server) his partner will perform the same task for a different set of files. The player then makes a judgment regarding the partner's performance. Between rounds, the player has the option of switching partners, and must explain the reasoning behind this decision. This choice serves as the decision point to elicit the fundamental attribution error. If the player switches partners and attributes it to a poor strategy on the part of the partner, even though the partner used a good strategy, then they are committing the fundamental attribution error. The focus of this LO is to train the player to recognize the difference between personal factors (internal to the entity, such as ability or personality) and situational factors (external to the entity, such as attacks on the datadroids) that may impact the partner's performance. In this learning opportunity, between 30% and 80% of the participants switch partners due to a fundamental attribution error. It should be noted that players switched partners more often, but for other reasons (e.g., liked their costume, etc).

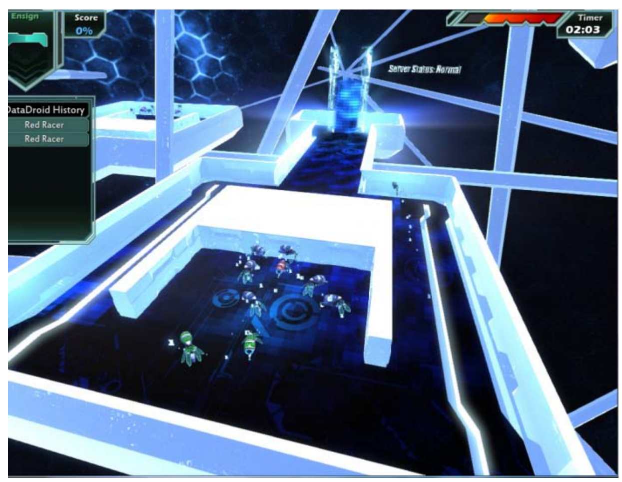

Fig. 2. Evaluating teamwork in fundamental attribution error

C. *Investigate Learning Opportunity: Making Sense of Incomplete and Ambiguous Information*

After a crisis occurs on the space station, the Evaluator conducts personnel interviews to determine the root cause of the emergency. This action phase LO exposes the player to information, provided incrementally (in three episodes separated by other LOs), explaining the potential causes of the crisis. Each opportunity presents three pieces of information, one that confirms a particular cause, one that disconfirms a cause, and one that provides a conflicted explanation of a potential cause. At three different times over the course of the action phase, the player must evaluate each piece of information and make a determination about the most likely cause of the crisis. This LO provides an opportunity to expose fundamental attribution error. In our experiments, the fundamental attribution error occurs in 30% of the initial assessments in this LO, and decreases over time.

D. Real-time Strategy Learning Opportunites

We created a suite of 11 real time strategy LOs (Fig. 4) that were all very similar in their game mechanics: five for Fundamental Attribution Error, three for Confirmation Bias, and three for Bias Blind Spot. These LOs allow players to observe a potential bias situation manifesting in another crew member and to fix it. These LOs were designed around a central monitor and repair game mechanic. Participants needed to monitor NPC thoughts and identify potential problems in order to complete the needed repairs to the space station.

The goal of each of these tasks is to complete repairs to the space station, allowing the disabled systems to activate and the crew to evacuate. In each of the 11 LOs, the player is able to see a birds-eye view of a specific section of the space station, with rooms, corridors, and crew members. In each section, crew members are trapped due to damage caused by the crisis. The player's goal is to order the trapped crew members to complete repairs of nearby generators and doors in order to allow everyone to escape. Meanwhile, the player must also monitor the crew members' thoughts for bias. Crew members with potentially biased thoughts will disrupt repair progress, so the player must flag the thought to alert the crew member and prevent bias from occurring. The player can click on each crew member to display a details screen that lists their attributes and current thoughts; participants also give orders and report thoughts from this screen. If the player reports a thought that is not biased, the crewmember will complain and the Evaluator will explain that the thought is not biased.

Fig. 3. Real-time strategy learning opportunity screen shot

To support the game narrative, androids never have biased thoughts and instead report their log output. Human thoughts can be biased or unbiased, and the thought library includes distractor thoughts that have nothing to do with biases, but support training and engagement.

III. GAME DEVELOPMENT CYCLES

Development of the serious game Heuristica was science driven, incorporating learning theories from cognitive psychology, game research, multimedia learning, and the debiasing literature (Fig. 4). We conducted interviews with expert analysts, who spend their days evaluating and assessing information in situations in which these cognitive biases could emerge. Together, these data were used to design ideas for training participants to mitigate biases.

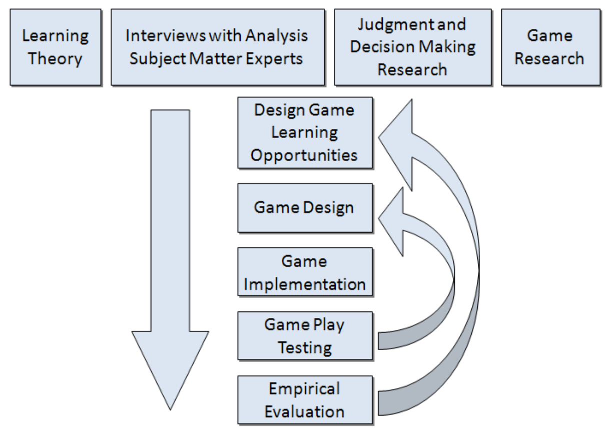

Fig. 4. The iterative design cycle of Heuristica.

Game development included three cycles, each lasting three to four months, to develop significantly different versions of the game. We used a theory and instructional design driven approach to develop Heuristica, leveraging theories of learning applied to tasks involving analysis, to support that learning in the game design. We built a variety of LOs to support better learning retention and transfer.

LOs were generated and iteratively improved throughout all of the testing cycles. Fig. 4 gives an overview of the general process. These LOs generally resulted from a literature review and discussion of the concepts to be learned. From these inputs, we generated guiding documents and outlines to keep concepts aligned as the task and accompanying narrative were built. Once a fully formed game design document was completed, the team was briefed in a presentation pulling together the scientific background, game mechanics, and task flow. This feedback helped serve as a check on face validity.

For most of the LOs, we created paper prototypes using Microsoft Power Point for initial playtesting before the decision making and game ideas were implemented in 3D. While not all aspects of the game could be recreated in Power Point, we tested the text and basic approach. For example, we observed play testers searching for information in a paper version of the emergency medical Diagnose LO. This enabled us to quickly respond to any problems related to instructions or participants' task understanding.

In parallel with this step, art and storyboard, and interface mockups were created. Once paper play testing results were as intended, the LO was instantiated into 3D game form, followed by play testing of the LO in the game. Each of these stages represents at least one pass through the Fig. 4 cycle. Once arriving at a fully functional environment, the cycle continues as the LO is tested in context of the overall game. The game was play tested regularly during this stage, and we often used cognitive task analysis techniques, such as verbal protocols, to unpack what players were thinking as they played [14]. This iterative process yields high quality results and reduces development costs for game programming and art requirements.

The development team has continually iterated on the game's design and graphics, with the most recent updates focusing on the action phase of the game. Effects such as smoking panels, sparking wires, and damaged walls serve to complement the darker emergency lighting about the space station. The game environment has been designed to be visually striking and engaging to encourage players to play for longer durations and to repeat play if given the option. We also implemented a navigation guide to help the player navigate the space station, and to reach the location of the next LO. The game's graphical user interface and other on screen elements received a complete update to improve readability and match the futuristic narrative.

Finally, the dialog for the Evaluator and all other NPCs was recorded by professional voice actors and their voice over was added to the game. The voice-overs were redundant with the text on the screen. This enhanced realism and supported better understanding and retention of game information, such as instructions and feedback provided by the Evaluator.

Based on feedback from play testers and the Heuristica team, we have continually revised and implemented new LOs and features. The real time strategy LOs were designed to teach bias recognition by presenting the thoughts of NPCs. These newer LOs completed our previous two-part LOs, which aimed to elicit bias in the player's own thoughts and actions, and train them to recognize that it was happening.

For the empirical evaluation of the game and to support its use as an experimental platform for game research, Heuristica is fully instrumented to capture a number of measures such as specific player choices and behaviors, response times, and scores. The game syncs these data to a central server from multiple physical sites.

IV. Learning Features in the Game

In designing Heuristica, we incorporated several game variables which have been identified in the cognitive literature as critical for improving learning. Three of these elements are reward, real time feedback, and a tailored game play experience (i.e., "intelligent tutoring"). These features have been iteratively improved over the three major development cycles.

A. Reward

We explored the effects of reward outcomes to player performance on learning by manipulating the amount of points won during LOs (high vs. low) and when they were delivered (continuously or all at once) to see what impact these variations would have on training success. Certain LOs would time-out after a few minutes of play, thereby penalizing the player. This was done to support learning and engagement.

B. Real-time Feedback

The current game version includes implementations of Real-time Feedback features such as head up display effects, sounds, and trial feedback and post-LO feedback.

Real-Time Feedback has been shown to improve player performance [2]. Commercial games use a variety of strategies to provide immediate and effective feedback to the player, including sound effects signaling player accomplishments, summary information about the state of the game or performance, individual trial feedback, and score updating. Our current game version includes the following kinds of immediate feedback that players may receive: cognitive bias mitigation feedback after each trial that explains why the player's actions were biased or unbiased; feedback about how the player's actions do or do not help the player complete the mission (if playing an RTS LO); points added immediately on the HUD (see Fig. 3), which can eventually lead to a rank increase; after action performance reports (summary feedback screen) at the end of each LO; and on-screen splash messages updating player progress after each trial (Fig. 5).

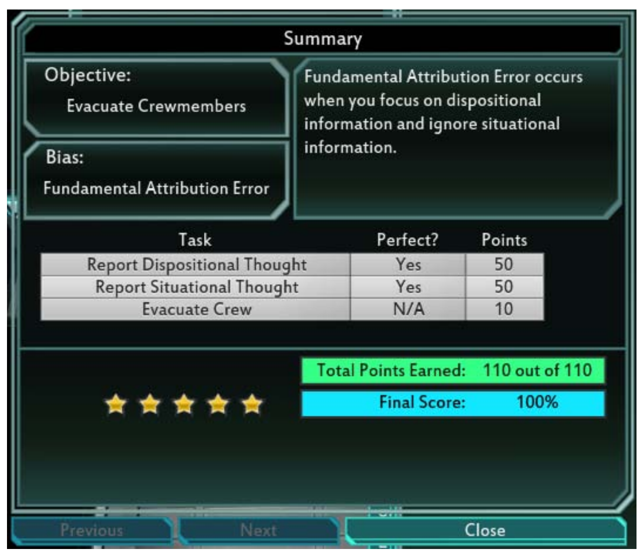

Fig. 5. Head up display with post learning opportuntiy reporting

C. Intelligent Tutoring: The Student Model

The game play is controlled by a computational model that captures each student's play, called the Student

Model. It is a computational model that dynamically modifies the LO order based on player initial knowledge and the player's in-game performance. A discussion of this aspect of the game and its effectiveness is presented in a separate paper at this conference [4].

The Student Model, using an intelligent tutoring systems approach, tailors the game experience to individual player performance. Based on the player's previous performance, the Student Model determines which LO a player should encounter next using their current curriculum model derived from the cognitive bias learning objectives. Each LO in the game is mapped to specific learning objectives, such as "what is confirmation bias," "how does fundamental attribution error manifest," or "what are mitigation strategies for bias blind spot." The Student Model tracks the activity of a participant while playing Heuristica, using game behavior and final LO scores to evaluate what the student had learned and still needs to improve upon. Next, the intelligent tutor auto-updates within the game to represent the current state of a participant's knowledge and select the subsequent set of LOs for presentation. For more information on this piece of software and how it manifests in Heuristica, please consult [4]. Together this individual customization allows the players to focus their effort on activities that support the training that they need, which can lead to faster and more effective training.

V. Conclusions

Our experience during the design, iterative development, and evaluation of Heuristica thus far has indicated that serious games can be effective. One challenge can be coupling good game design with good science, but the payoff can be large. Both playtesting and empirical evaluations played important roles in our process and provided the interdisciplinary team with critical information and feedback. Serious games have the potential for cultivating a variety of learning experiences in an environment where failure is allowed to occur. We found that players are able to reach goals and learn a system more effectively when they have real-time feedback that provides information about not only players' outcomes, but also their strategies. As such, serious games can be effective training tools. However, not until more empirical research is done to quantify the effects of various game features on learning will we be able to fully realize their potential and improve the design of video games for training. This paper describes one contribution to that effort. A variety of academic and industry research labs are focusing on this problem, and more work remains to be done in the future.

Acknowledgment

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory contract number FA8650-11-C-7177. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: 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.

We would also like to thank the Virtual Heroes art team, programming team, and quality control team who implemented and tested Heuristica. We would also like to thank the play testers who made Heuristica better and the reviewers who provided useful feedback that improved this paper.

References

[1] Gee, J. P. (2003). "What video games have to teach us about learning and literacy". New York: Palgrave Macmillan.

[2] Stone, R. (2005). "Serious Gaming–Virtual Reality's Saviour. Proceedings of Virtual Systems and MultiMedia annual conference", VSMM (pp. 773–786).

[3] Veinott, E., et al. (2013, September) The effect of 3rd person persepctive and session duration on training decision making in a serious video game. Paper to be presented at IEEE Games Innovation Conference (IGIC), Vancouver, BC.

[4] Whitaker, E. Trewitt, E., Holtsinger, M., Hale, C., Veinott, E., Argenta, C., and Catrambone, R. (2013, September). The effectiveness of intelligent tutoring on training in a video game. Paper to be presented at IEEE Games Innovation Conference (IGIC), Vancouver, BC.

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

[6] Lehner, P., Adelman, L., Cheikes, B., Brown, M., (2004). Confirmation bias in complex analysis. *IEEE Transactions on systems, man and cybernetics*, *38*, 584-592.

[7] Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, *20*(3), 273–281.

[8] Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In Berkowitz, L. (Ed.) *Advances in experimental social psychology, Vol. 10*. New York: Academic Press, 173-220

[9] Schwarz, N.; Sanna, L.; Skurnik, I.; and Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, 39, 127-161.

[10] Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 111*, 781–799.

[11] A. Lang, "Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages," *Journal of Communication*, vol. 56, no. 1, pp. 57–80, 2006.

[12] Crandall, B., Klein, G., and Hoffman, R. (2006) *Working Minds*. Bradford Books: NY.

[13] Mersch, E., Fox, O., Flach, J., Bennett, K., Leonard, J., & Holt, J. (2013, September) "Mitigation of the confirmation bias using a game-based trainer". Poster to be presented at the Human Factors and Ergonomics Society International Annual Meeting, San Diego, CA

[14] Yee, N. (2006). Motivations for Play in Online Games. *CyberPsychology & Behavior. December 2006,* 9 (6), 772-775.