Exploring the Impact of Role Model Avatars on Game Experience in Educational Games

Dominic Kao
MIT
77 Massachusetts Ave
Cambridge, MA 02142 USA
dkao@mit.edu

D. Fox Harrell
MIT
77 Massachusetts Ave
Cambridge, MA 02142 USA
fox.harrell@mit.edu

CHI PLAY 2015, October 03–07, 2015, London, United Kingdom.
ACM 978-1-4503-3466-2/15/10.
http://dx.doi.org/10.1145/2793107.2810291

Abstract
Studies show that role models can boost academic performance [11, 9, 8]. In this paper, we describe an experiment (N = 890) exploring the use of (a) scientist role models, (b) athlete role models and (c) simple geometric shapes, as game avatars. Using the Game Experience Questionnaire (GEQ) [3], we find that over all participants, the use of avatars that looked like scientist and athlete role models led to highest flow and immersion. For female participants, the use of scientist avatars led to highest immersion and positive affect, and lowest tension and negative affect. The results here indicate that role model avatars have the potential to positively affect player game experience. This may especially be impactful for educational games, in which higher engagement could in turn influence learning outcomes [1].

Author Keywords
Avatars; Educational Games; Virtual Identity

ACM Classification Keywords
K.8.0 [**Personal Computing**]: General – Games

Introduction
Stereotype threat, the theory that the mere *idea* of conforming to a stereotype can hinder one's performance, is well-studied in the social sciences [15]. In one such

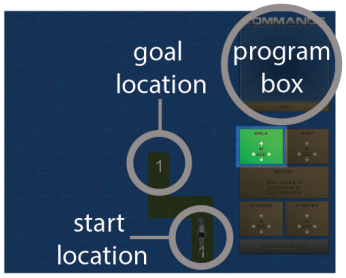

Figure 1: Level 1 in *Mazzy* introduces the basic game mechanics.

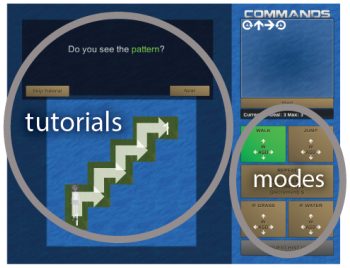

Figure 2: Level 6 introduces looping.

study, men and women college students were invited to watch a set of six television commercials. For half the participants, two of the commercials depicted women in gender-stereotypical ways. For the remaining half, there was no gender content in the commercials. All the participants were then asked to help another student with problems in mathematics. The results showed that women that had seen the commercials depicting women in stereotypical ways chose fewer math problems, performed worse on the ones they did choose, and reported being less interested in math-related college majors and careers [16]. Stereotype threat (which is active even without explicit cues like stereotypical images) could have possible implications for games via virtual avatars; recent studies have suggested that this phenomenon could impact performance and engagement in educational games [6, 4].

Researchers have studied many approaches on how to mitigate against stereotype threat. One successful approach has been through providing people with role models. McIntyre et. al showed that simply reading essays about successful women before an exam, can alleviate the negative effects of stereotype threat [14, 13]. In this paper, we study how role model *avatars* impact player performance and game experience in an educational game. We use responses to the Game Experience Questionnaire (GEQ) [3] to measure the following dimensions of game experience: (a) flow, (b) immersion, (c) competence, (d) challenge, (e) positive affect, (f) negative affect and (g) tension. We are particularly interested in how female participants react to the role model avatars, since the social sciences literature has shown that role models (in particular STEM role models) can positively impact women's self-esteem and academic test performance [7, 11, 13, 14].

The Game

The experiment takes place in a STEM learning game called *Mazzy* [5]. *Mazzy* is a game in which players solve mazes by creating short computer programs. In total, there are 12 levels in this version of *Mazzy*. Levels 1-5 require only basic commands. Levels 6-9 require using loops. Levels 10-12 require using all preceding commands in addition to conditionals. See Figures 1 and 2. *Mazzy* has been used previously as an experimental testbed for evaluating the impacts of avatar type on performance and engagement in an educational game [6, 4].

Methods

Our experiment aims to compare three avatar types: (a) scientist role models, (b) athlete role models, and (c) simple geometric shapes. The goal is to see if participants of different avatar type have differing game performance and game experience as measured by the GEQ. We strongly suspected ahead of time that (1) scientist avatars would outperform athlete and shape avatars, and (2) athlete avatars would outperform shape avatars.

Avatar Conditions

The three avatar conditions we tested were:

a. Scientist Avatars
b. Athlete Avatars
c. Shape Avatars

In each condition, players selected (inside the game) from a pool of eight possible choices. The pool of role models is composed of famous individuals, selected for diversity (i.e., exactly half the role models are female, and exactly half the role models are black or African American). When a user selects an avatar, there is a three-sentence summary presented of the avatar (e.g., "You've selected Albert Einstein. Albert Einstein was a German-born

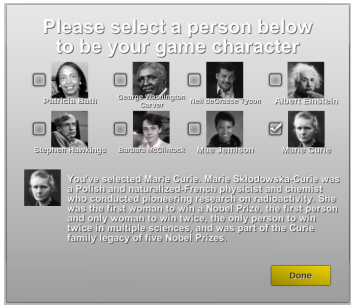

Figure 3: Scientist avatar.

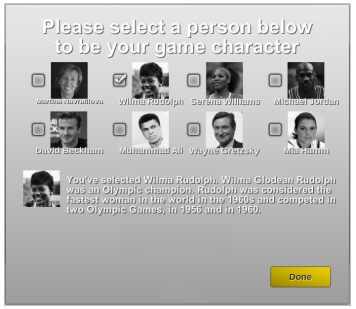

Figure 4: Athlete avatar.

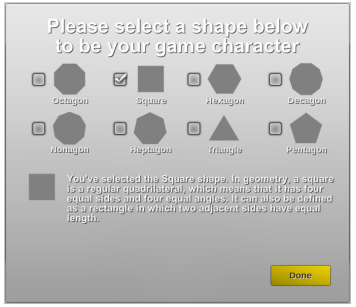

Figure 5: Shape avatar.

theoretical physicist., etc.). These were taken verbatim from their Wikipedia article. Avatars are always presented in a randomized ordering on the screen (see Figures 3, 4 and 5). Inside the game, the avatar consists of a 60 x 60 pixel game character that moves according to the user's programs.

Quantitative and Qualitative Measures

For performance, we only analyze the number of levels completed by players. For measuring game experience, we use the GEQ [3]. We also included a single, 5-item Likert scale question on how the user felt towards the game character (1:Strongly Negative to 5:Strongly Positive).

Participants

890 participants were recruited through Mechanical Turk. The data set consisted of 528 male, and 362 female participants. There were 712 white participants, 61 black or African American, 29 Chinese, 15 Filipino, 9 Korean, 8 American Indian, 8 Asian Indian, 8 Vietnamese, 8 Other Asian, 3 Japanese, 1 Native Hawaiian, 1 Guamanian or Chamorro, and 27 other. Participants were between the ages of 18 and 75 (M = 31.4, SD = 9.0), and were all from the United States. Participants were reimbursed $1.50 to participate in this experiment.

Design

A between-subjects design was used: avatar type was the between-subject factor. Participants were randomly assigned to a condition.

Protocol

Prior to starting the game, players were informed that they could exit the game *at any time* via a red button in the corner of the screen. When participants were done playing (either by exiting early, or by finishing all 12 levels), participants returned to the experiment instructions, which then prompted them with the GEQ and then a demographics survey.

Results

Aggregate

17 of the participants completed 0 levels. 3 of these were in the athlete condition, 4 in the scientist condition, and 10 in the shape condition. Being that the very first level of the game requires only following a set of simple directions (a basic tutorial level), these participants invested minimal effort and provide data of limited use. These participants are therefore excluded from further analysis. We report on data from 873 participants. The one-way ANOVA found no significant effect of avatar type on levels completed, $F(2, 870) = 0.89$, $p = 0.41$ (see Table 1).

Table 1: Overall level completion statistics.

Avatar Condition	N	Mean	SD
Scientist	278	7.47	3.02
Athlete	308	7.40	2.87
Shape	287	7.16	2.84

A MANOVA revealed a statistically significant difference in GEQ responses and avatar ratings based on the participant's avatar type, $F(86, 1654) = 2.23$, $p < .0001$; Wilk's $\lambda = 0.803$, partial $\eta^2 = .10$. See Figure 6. Univariate testing found the effect to be significant for the following items: "I was deeply concentrating on the game" (*flow*), $F(2, 870) = 5.73$, $p < 0.005$; "I was fully occupied with the game" (*flow*), $F(2, 870) = 3.36$, $p < 0.05$; "It was aesthetically pleasing" (*immersion*), $F(2, 870) = 5.63$, $p < 0.005$; and "I felt imaginative" (*immersion*), $F(2, 870) = 4.30$, $p < 0.05$. Avatar rating was also found to differ between conditions, $F(2, 870) = 5204$, $p < .0001$.

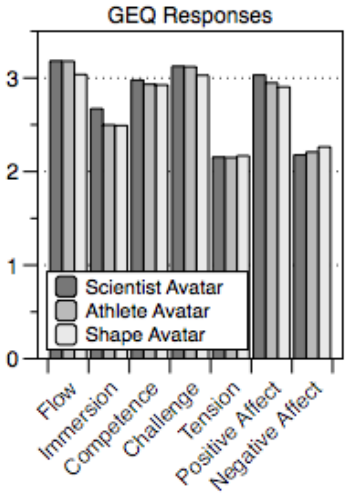

Figure 6: Game Experience Questionnaire (GEQ) responses for all participants.

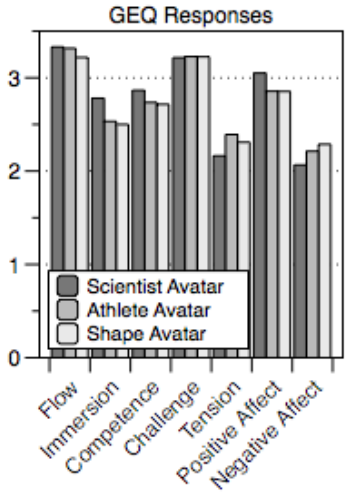

Figure 7: Game Experience Questionnaire (GEQ) responses for female participants.

In order to compare the effects of avatar type on these measures, we additionally calculated posthoc comparisons (LSD) between all conditions. The pair-wise comparisons revealed that the scientist condition GEQ rating was higher on "It was aesthetically pleasing" (*immersion*), $p < 0.05$, and "I felt imaginative" (*immersion*), $p < 0.005$, than the athlete condition. The scientist condition was also higher on "I was fully occupied with the game" (*flow*), $p < 0.05$, and "It was aesthetically pleasing" (*immersion*), $p < 0.005$, than the shape condition. The athlete condition GEQ rating was higher on "I was deeply concentrating on the game" (*flow*), $p < 0.0001$, and "I was fully occupied with the game" (*flow*), $p < 0.05$, than the shape condition. The scientist avatar rating was higher, $p < 0.0001$, than the athlete and shape condition. The athlete avatar rating was higher, $p < 0.0001$, than the shape condition.

Female Participants

The one-way ANOVA found no significant effect of avatar type on levels completed by female participants, $F(2, 355) = 1.30$, $p = 0.27$ (see Table 2).

Table 2: Female participant level completion statistics.

Avatar Condition	N	Mean	SD
Scientist	108	7.57	3.15
Athlete	122	7.00	3.00
Shape	128	7.02	2.93

The MANOVA revealed a statistically significant difference in GEQ responses and avatar ratings based on the participant's avatar type, $F(86, 626) = 1.31$, $p < .05$; Wilk's $\lambda = 0.718$, partial $\eta^2 = .15$. See Figure 7. Univariate testing found the effect to be significant for the following items: "I was interested in the game's story" (*immersion*), $F(2, 355) = 3.87$, $p < 0.05$; "It was aesthetically pleasing" (*immersion*), $F(2, 355) = 3.41$, $p < 0.05$; "I felt imaginative" (*immersion*), $F(2, 355) = 3.38$, $p < 0.05$; "I felt irritable" (*tension*), $F(2, 355) = 4.95$, $p < 0.01$; "I could laugh about it" (*positive affect*), $F(2, 355) = 3.28$, $p < 0.05$; and "I was distracted" (*negative affect*), $F(2, 355) = 3.34$, $p < 0.05$. Avatar rating was also found to differ between conditions for female participants, $F(2, 355) = 19.49$, $p < .0001$.

In order to compare the effects of avatar type on these measures, we again additionally calculated posthoc comparisons (LSD) between all conditions. The pair-wise comparisons revealed that the scientist condition was higher on "I felt imaginative" (*immersion*), $p < 0.05$, and "I could laugh about it" (*positive affect*), $p < 0.05$, and lower on "I felt irritable" (*tension*), $p < 0.005$, and on "I was distracted" (*negative affect*), $p < 0.05$ than the athlete condition. The scientist condition was higher on "I was interested in the game's story" (*immersion*), $p < 0.01$, "It was aesthetically pleasing" (*immersion*), $p < 0.01$, and "I could laugh about it" (*positive affect*), $p < 0.05$, and lower on "I felt irritable" (*tension*), $p < 0.05$, and "I was distracted" (*negative affect*), $p < 0.05$, than the shape condition. There were no significant differences between the athlete and shape conditions. The scientist avatar rating was higher, $p < 0.0005$, than the athlete and shape conditions. The athlete avatar rating was higher, $p < 0.05$, than the shape condition.

Discussion

In the results, we have seen that using scientist avatars resulted in the highest scores on *immersion*, and that scientist and athlete avatars resulted in the highest scores on *flow*. For female participants, scientist avatars resulted in the highest scores on *immersion*, and *positive affect*,

and in the lowest scores on *tension* and *negative affect*. We observe that there is consistent ordering in that scientist avatars are better than athlete avatars, and athlete avatars are better than shape avatars, across all subjective measures.

These results corroborate the findings in the social sciences, and demonstrate that those same findings are likely applicable via avatars inside educational games. For example, there have previously been three criteria for determining the effectiveness of a role model for boosting academic performance. (1) First, the role model should be perceived as competent [12]. Clearly, both the scientist and athlete avatars fulfill this condition. (2) Second, the role model should be perceived as an ingroup member, for instance the same gender or race [7, 14, 9]. This is also made possible for female and African American participants by having selected for diverse role models. (3) Third, the role model's record of success in domains where the role model's group is negatively stereotyped should be readily available [10, 2, 11]. To faciliate this, we have used only role models that are well known, and also provide a paragraph describing their achievements. Given that the experimental setting is a STEM education game, it is not of surprise that we have found scientist role models to be the more effective avatar type (e.g., [11]).

Conclusion

In conclusion, the implications of this work for educational games is that certain avatar could be more effective than existing avatar types in educational games. Role model avatars may be a way of improving the game experience for individuals. This is especially true for those participants that come from negatively stereotyped social categories in STEM (e.g., women, black or African American, Latino/a, etc.). We hope to continue data collection and be able to test these findings across more social categories.

Acknowledgements

We thank the anonymous reviewers for their feedback. This work is supported by National Science Foundation (NSF) Award IIS-1064495, a QCRI-CSAIL Collaboration Grant, and a Natural Sciences and Engineering Research Council of Canada (NSERC) fellowship.

References

[1] Blumenfeld, P. C., Kempler, T. M., and Krajcik, J. S. Motivation and Cognitive Engagement in Learning Environments. In *The Cambridge Handbook of the Learning Sciences*. 2005, 475–488.

[2] Buunk, A. P., Peiró, J. M., and Griffioen, C. A positive role model may stimulate career-oriented behavior. *Journal of Applied Social Psychology 37* (2007), 1489–1500.

[3] IJsselsteijn, W., De Kort, Y., Poels, K., Jurgelionis, A., and Bellotti, F. Characterising and Measuring User Experiences in Digital Games. *International Conference on Advances in Computer Entertainment Technology 620* (2007), 1–4.

[4] Kao, D., and Harrell, D. F. Exploring construction, play, use of virtual identities in STEM learning. *Jean Piaget Society Annual Conference* (2015).

[5] Kao, D., and Harrell, D. F. Mazzy: A STEM Learning Game. *Foundations of Digital Games* (2015).

[6] Kao, D., and Harrell, D. F. Toward Evaluating the Impacts of Virtual Identities on STEM Learning. *Foundations of Digital Games* (2015).

[7] Lockwood, P. "Someone like me can be successful": Do college students need same-gender role models? *Psychology of Women Quarterly 30*, 1 (Mar. 2006),

36–46.

[8] Lockwood, P., and Kunda, Z. Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology 73*, 1 (1997), 91–103.

[9] Marx, D. M., and Goff, P. A. Clearing the air: the effect of experimenter race on target's test performance and subjective experience. *The British journal of social psychology / the British Psychological Society 44* (2005), 645–657.

[10] Marx, D. M., Ko, S. J., and Friedman, R. a. The Obama Effect: How a salient role model reduces race-based performance differences. *Journal of Experimental Social Psychology 45*, 4 (July 2009), 953–956.

[11] Marx, D. M., and Roman, J. S. Female Role Models: Protecting Women's Math Test Performance. *Personality and Social Psychology Bulletin 28* (2002), 1183–1193.

[12] Marx, D. M., Stapel, D. a., and Muller, D. We can do it: the interplay of construal orientation and social comparisons under threat. *Journal of personality and social psychology 88*, 3 (2005), 432–446.

[13] Mcintyre, R. B., Lord, C. G., Gresky, D. M., Eyck, L. L. T., and Bond, C. F. A Social Impact Trend in the Effects of Role Models on Alleviating Women's Mathematics Stereotype Threat. *Current Research in Social Psychology 10*, 9 (2005), 1–26.

[14] McIntyre, R. B., Paulson, R. M., and Lord, C. G. Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology 39*, 1 (2003), 83–90.

[15] Steele, C., and Aronson, J. Stereotype Threat and the Intellectual Test Performance of African Americans. *Journal of personality and social psychology* (1995).

[16] Steele, C. M. Whistling Vivaldi and other clues to how stereotypes affect us. In *Whistling Vivaldi*. 2010.