

Exploring the Use of Role Model Avatars in Educational Games

Dominic Kao and D. Fox Harrell

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract

Research has indicated that role models have the potential to boost academic performance (Marx and Roman 2002; Marx and Goff 2005). In this paper, we describe an experiment exploring role models as game avatars in an educational game. Of particular interest are the effects of these avatars on players' performance and engagement. Participants were randomly assigned to a condition: a) user selected role model avatar, or b) user selected shape avatar. Results suggest that role models are heavily preferred. African American participants had higher game affect in the role model condition. South Asian participants had higher self-reported engagement in the role model condition. Participants that completed ≤ 1 levels had higher performance in the role model condition. General trends suggest that the role model's gender and racial closeness with the player, could play a role in player performance and self-reported engagement as consistent with the social science literature.

Introduction

It is well known that the underrepresentation of ethnic minority groups and women represents a huge untapped well of innovative, productive contributors to the nation's Science, Technology, Engineering and Mathematics (STEM) fields and workforce. Indeed, U.S. Educational Secretary Arne Duncan has said that addressing this problem "is not just an economic imperative, but a moral imperative." (Block and Siegel 2009). Moreover, research has shown that promoting STEM among learners from underrepresented groups may achieve more than turning out a greater *number* of engineers it can produce a greater *diversity of ideas* (Genskow 2006; Cox and Blake 1991).

Role model avatars may be able to promote STEM in educational games. It has been shown that role models provide an academic boost for African American (Blanton, Crocker, and Miller 2000) and female (Marx and Roman 2002) participants. The purpose of the work here is exploring the impact of using a role model as game avatar versus using a shape as game avatar. Figures 1 and 2 illustrate the avatars



Figure 1: Player selected role model avatars.

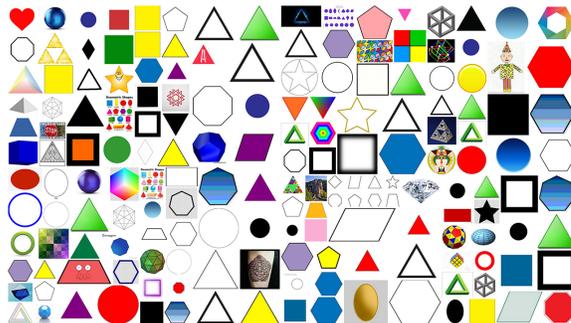


Figure 2: Player selected geometric shape avatars.

chosen by players. Previous research has studied a similar dichotomy between avatars that are anthropomorphic and non-anthropomorphic (Morie and Verhulsdonck 2008; Waytz, Heafner, and Epley 2014; Kao and Harrell 2015c; 2015a; Perlin 2002).

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 (Steele 1997). Recent work suggests stereotype threat can be activated by avatars in a STEM learning game, resulting in lower self-reported engagement (Kao and Harrell 2015a; 2015d). Studies have demonstrated that role models are effective at mitigating stereotype threat (Blanton, Crocker, and Miller 2000; Marx and Roman 2002; Marx and Goff 2005). However, no stud-

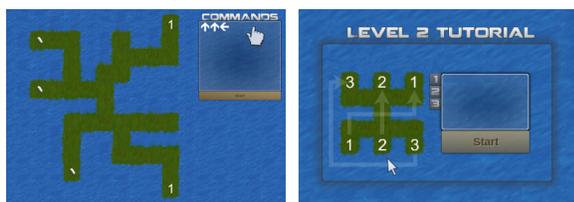


Figure 3: In *Mazzy*, players write “code” to navigate a maze.

ies on using role model avatars exist to date.

Here, we describe an exploratory study (N = 357) comparing user role models to user geometric shape avatars. The experiment was run in a STEM education game. Our analyses suggest that role models are the preferred avatar. African American participants had higher game affect in the role model condition. South Asian participants had higher engagement in the role model condition. Participants that completed ≤ 1 levels had higher performance in the role model condition.

The paper is structured as follows: Section 2 describes the STEM game that is the experimental testbed. Section 3 describes work on stereotypes, role models, and virtual identities. Section 4 details the methods used. Section 5 reports the results. Section 6 has our discussion of the results. We make concluding remarks in Section 7 and discuss about future work in Section 8.

The Game

The game we used is *Mazzy*; it is a STEM learning game designed to be fun, but also to foster computational thinking. *Mazzy* has been used as an experimental testbed for evaluating the impacts of avatar type on performance and engagement (Kao and Harrell 2015a; 2015d). Players use the keyboard to write procedures to control a character (see Figure 3) and guide it to the end.

Mazzy’s design is grounded in an influential pedagogical approach called “constructionism,” in which building objects is central to the process of learning (Papert and Harel 1991). Constructionism originates in the principle that learning a new concept or idea is easier if it can be assimilated into existing models (Piaget 1976). In *Mazzy*, the character is “body syntonic” (Papert 1993); this means players can identify with it and its motion in space. Players in *Mazzy* use code blocks, procedural thinking, looping, conditional statements, etc. (Kao and Harrell 2015b). Players are learning computing by creating programs via a real concrete object that can be manipulated.

Mazzy uses symbols to represent code instead of natural language. This has the advantage of being very simple, since the notion of making a syntactical mistake such as forgetting a closing bracket or misspelling a command is impossible. This is also to make code understandable since symbols represent their purpose. When players run a program, each symbol is highlighted as it is processed (a form of “debugging”). This stems from the philosophy that coding is an iterative process and that mistakes are natural.

In this version of *Mazzy* used for this study (the system is highly customizable and extensible), there are three levels. Adding new levels is not technically difficult, although developing levels that are both fun and effective for learning requires skillful game design. Each of the levels has an animated tutorial to help teach the player. Players can pickup bonus items inside each level, but require more complex programs. Levels get harder and introduce new mechanics. Level two requires the player to program multiple characters in parallel, and level three requires the player to program boolean logic into the level map. On average about 1/3 of participants complete the game.

Related Work

Stereotype Threat

Stereotype threat is the phenomenon of being at risk of confirming, as self-characteristic, a negative stereotype about one’s group (Steele and Aronson 1995). This phenomenon motivates some of our efforts investigating the impacts of avatars on players. We are motivated by the fact that users’ representations may act as triggers prompting more positive or negative outcomes depending on the social group of the user as visually represented by the avatar’s appearance. In other words, we are building toward addressing the impact of avatars on stereotype threat. Stereotype threat can lead to a number of harmful consequences, ranging from decreased performance (e.g., women performing worse in math when their female identity is made salient) (Shih, Pittinsky, and Ambady 1999) to altered professional aspirations (e.g., stereotype threat undermines sense of belonging and reduces women’s desire to pursue math in the future) (Good, Rattan, and Dweck 2012).

Techniques such as deemphasizing threatened identities (Stricker and Ward 2004) and endorsing an incremental view of intelligence (Aronson, Fried, and Good 2002) have been seen to reduce, and in some cases eliminate, stereotype threat. Recent works suggests that stereotype threat persists in virtual environments (Kao and Harrell 2015a; 2015d). This is consistent with the observation of stereotype-related phenomena identified by other researchers, e.g., Yee demonstrated in (Yee et al. 2011) that often players’ behaviors conform to stereotypes associated with their avatar’s gender. In this work, our systematic study of the impacts of virtual identities on learners will enable us to develop systems that could help in inoculating users against stereotype threat in STEM learning.

Role Models

The term “role model” comes from Merton. He hypothesized that an individual compares themselves to references (other people) that occupy a desirable standing to which the individual aspires (Merton 1936). Effective use of these role models has been used to reduce stereotype threat. In one such study, participants read anywhere between 0-4 biographies of successful women. All the participants then took a difficult math test. The female participants that did not read any biographies performed worse than men. However, the more biographies that female participants read,

the better they performed. Those female participants that had read 4 biographies performed at the same level on the math test as the men (McIntyre, Paulson, and Lord 2003; McIntyre et al. 2005).

There are three factors that can increase a role model's effectiveness. The first is the perception of the role model as competent (Marx, Stapel, and Muller 2005). The second is sharing common attributes such as gender or race, since they are seen as an ingroup member that has overcome stereotypes (Lockwood 2006; Marx and Roman 2002). The third is that the role model should have achieved success (Buunk, Peiró, and Griffioen 2007). Enforcing these criteria is non-trivial. Finding a role model of the same race and gender as users requires soliciting demographics. However, soliciting these details prior to the task is itself a possible activation of stereotype threat (Shih, Pittinsky, and Ambady 1999). Instead, we designed the experiment such that players could choose *any* role model. Players (often) naturally fulfill these criteria, this also affords analyses on the criteria themselves, e.g., participants that selected same race versus different race avatars.

Virtual Identities

Harrell describes digital self-representations as selective projections of some aspects of a real player (e.g., preferences, control, appearance, personality, understanding of social categories, etc.) onto the actual implemented, virtual, representation (Harrell 2013). As such, Harrell's notion of a "blended identity" is an approach based on looking at structural mappings from one domain to another that is central to the understanding of virtual identities in this project (Harrell 2010). This concept builds upon James Gee's notion of the "projective identity", which can be described as "manifesting the ways that real player values are reconciled with values understood as being associated with avatars." (Gee 2007; Harrell et al. 2014). Relating in-game behavior to real-world identities, such as demographic segments (Lim and Harrell 2013; 2014) has demonstrated useful insight into understanding how to match interaction mechanisms in digital media systems such as games to users in order to provide the most appropriate supports. Such supports can have strong impacts on user behaviors, such as has been shown by research on the "proteus effect", a phenomenon in which users conform to expected behaviors and attitudes associated with an avatar's appearance (Yee and Bailenson 2007).

Here, our focus is on matching avatar uses with supports for computer science learning by diverse players. To the best of our knowledge, there have been few studies of this nature that study the impact of avatar types on player engagement and performance. Linebarger et. al compared four avatar types on task performance in a virtual environment and concluded that "simpler, less computationally expensive avatar representations are quite adequate" (Linebarger and Kessler 2002). More recently, Domínguez et. al explored the impact of avatar color on performance in a virtual scavenger hunt (Domínguez and Roberts 2015). Previous work also suggests that certain avatar can prompt more negative dispositions towards the game (Kao and Harrell 2015c; 2015d).

Methods

The study we performed consisted of an experiment (N=357) inside of the game *Mazzy*. The study compares the impact of player selected role model avatars versus player selected shape avatars on player engagement and performance.

Avatar Conditions

The two avatar conditions we tested were:

1. Role Model: Avatar in the form of a role model.
2. Shape: Avatar in the form of a geometric shape.

Role model condition participants were asked to think of any type of role model (real or fictional). Shape participants were asked to think of any type of geometric shape. Participants were informed that whatever they came up with would be their game avatar. Participants then used Google image search to find an image representing their choice. This image was uploaded to the game and became the user's character that moved about the maze.

Task

The experimental task was to play *Mazzy*¹. There are three levels in this version of *Mazzy*. In the first level, players can click in the command box, after which they can use the arrow keys on the keyboard to input arrow commands. When participants click on the "start" button, the character begins to move according to the programmed arrows. The arrows are highlighted as each command is processed by the character. The character continues to move until either a) the character moves into a water tile, b) the character reaches the end of the maze, or c) the character has no more commands to process and has not reached the end. In case b), the player can advance to the next level. In any other case, the character disappears and the player should try again.

The second level is a direct extension of the first; players can now program three characters², all starting in different spots in the maze, and all having different goal locations. These start and corresponding end locations are marked with the same number. Beside the command box are three smaller buttons, clicking each of these brings up the code for each of the three characters. Clicking inside the command box allows players to modify the code for a single character. The same rules as in the first level apply to each of the three characters. In the second (and third) level, commands are highlighted for the character whose code is currently on-screen. The player may switch between each character's code view during execution.

The third level is similar to the second level in that there are three characters. However, all of them are already pre-programmed. The code can be viewed for all three characters, but their code cannot be modified. Two of the characters start at locations marked with an "x" (no corresponding end location), and one character starts at a location marked "1" (with a corresponding end location). The player can click *the*

¹<http://groups.csail.mit.edu/icelab/mazzy/>

²We note in the 2nd and 3rd levels there are multiple character copies. This *could* constitute a special case scenario.

map itself to toggle some tiles to either be water or grass. There are some combinations of toggles that allow only the character starting at “1” to reach the end location; doing so passes the level and the game. Toggling of these particular tiles can be done during code execution.

Bonus items are scattered in each level, which the players can optionally pick up. There are a total of nine bonus items, three in each level. In all levels, there is an animated tutorial in the top left of the screen demonstrating the mouse clicks and keyboard presses required to solve a simpler version of the current level; these levels have the same mechanics, only the mazes are reduced in complexity. Mousing over a help icon next to this animation provides a textual description of the goal for each level.

Quantitative and Qualitative Measures

The performance measures we recorded were:

- **Levels completed:** The number of levels completed.
- **Level attempts:** The number of attempts in each level.
- **Level bonus items:** Bonus items collected in each level.

The engagement measures we recorded were:

- **Enjoyment:** Enjoyment rating in each level.
- **Difficulty:** Difficulty rating in each level.

All subjective data was collected using a 5-point Likert scale. Players were also asked at the end of the experiment to rate how they felt overall with respect to the game, their progress, and their avatar, in addition to completing a demographics survey.

Participants

357 participants were recruited through Mechanical Turk. 129 of the participants were female. 193 of the participants were white, 112 south Asian, 22 black or African American, and the remaining participants divided among eleven other group categories. Participants were between the ages of 19-65 ($M = 31.4$) and were reimbursed \$2 to participate.

Design

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

Experiment Protocol

Prior to starting the task, players were told they could exit the game at any time. Then, for each condition players loaded the game in their web browser. After each level that players completed, players were presented with a screen showing the number of stars they had earned (corresponding to the number of bonus items they had collected); at this point in the procedure, players could either continue or replay the level. If they chose to replay the level, they were brought back to the previous level (with their previous code still intact). If they continued, they were then asked to report engagement (enjoyment and difficulty). When participants were done playing, they returned to the instructions, which prompted them with additional questions including the demographic survey.

Analysis

Data was extracted and imported into Statistical Package for Social Science (SPSS) version 22 for data analysis using multivariate analysis of variance (MANOVA). The dependent variables are- *total levels completed*, *total attempts*, *total bonus items*, *average enjoyment*, *average difficulty*, *avatar rating*, *progress rating*, *game rating*; and the independent variables are- *avatar type (role model vs. shape)*, *gender*, *race*. All the dependent variables are continuous variables. For the independent variables, both the avatar status (i.e., 0 = shape; 1 = role model) and gender (i.e., 0 = female; 1 = male) were dichotomous variables, and race (i.e., 1 = white, 2 = black or African American, 3 = south Asian, 4 = other) is a categorical variable. To detect the significant differences between user role model and user shape avatars, we utilized two-way or factorial MANOVA. The reason we chose factorial MANOVA is that we suspected that there would be an interaction effect between the independent variables. Also, we considered the variable-*age* as a covariate in the analysis (using MANCOVA), however, age was found not to be a significant covariate, as a result, it was not included in the subsequent analyses. First, we ran two-way MANOVA with avatar type and gender as independent variables, and then, another two-way MANOVA with avatar type and race. We also ran targeted independent-samples t-tests on the following groupings: *low performers (completed ≤ 1 levels)*, *high performers (completed ≥ 2 levels)*, *same vs. different gender role models*, and *same vs. different race role models*. These results are reported as significant when $p < 0.05$ (two-tailed). Before running MANOVAs, all the variables included in the analyses were checked, and there were no outlier detected. Prior to running our first MANOVA model, we checked both assumption of homogeneity of variance and homogeneity of covariance by the test of Levine’s Test of Equality of Error Variances and Box’s Test of Equality of Covariance Matrices; and both assumptions were met by the data ($p > .05$ for Levene’s Test, and $p > .05$ for Box’s Test). And for our second MANOVA model, using same tests, both assumptions were found not tenable ($p < .05$ for Levene’s Test except Average Enjoyment Rating and Levels completed, and $p < .05$ for Box’s Test). To address this violation issue, Pillai’s Trace value was considered instead of Wilk’s Lambda value.

Results

Avatar and Gender

Participants in the user role model condition had significantly higher avatar ratings compared to participants in the user shape condition. Our first MANOVA model that contained avatar types and gender as independent variables with a set of eight dependent variables was tested first. In this model, we looked for the main effect of avatar types, another main effect of gender, and an interaction effect of avatar types as well as gender on the set of dependent variables. The test results of the first MANOVA model indicated the main effect of avatar types as significant ($\lambda = .784$, $F(8, 244) = 8.399$, $p < .001$) whereas the main effect of gender and the interaction effect of avatar and gender were found

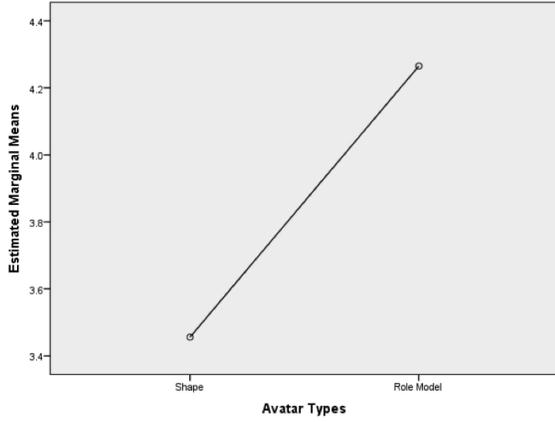


Figure 4: Avatar Ratings.

not be significant ($\lambda = .967$, $F(8, 244) = 1.53$, $p > .05$, and $\lambda = .986$, $F(8, 244) = .442$, $p > .05$). As a result, gender was removed from the first model, and the refined model was tested again. The refined first MANOVA model yielded a significant difference between user role model and user shape avatars on the dependent variables ($\lambda = .778$, $F(8, 248) = 8.841$, $p < .001$). Also, the tests of between subjects effects detected that avatar types are significantly different on avatar rating ($F(1, 255) = 59.97$, $\eta^2 = .19$, $p < .001$). Figure 4 shows that the participants who were in the user role model condition had higher avatar ratings compared to the participants who were in the user shape condition, and this difference is statistically significant.

Avatar and Race

There was a significant interaction between avatar type and race. African American participants had higher game affect, and marginally higher progress ratings in the role model condition. White participants had lower engagement in the role model condition. South Asian participants had higher engagement in the role model condition. In our second MANOVA model where avatar type and race were the independent variables with a set of eight dependent variables, the test result indicated an interaction effect of avatar and race on the set of dependent variables (Pillai's Trace = .165, $F(24, 726) = 1.76$, $p < .05$). Tests of between subject effects showed that the interaction effect of avatar type and race has a significant effect on the average enjoyment rating ($F(3, 247) = 4.05$, $\eta^2 = .05$, $p < .05$) and progress rating ($F(3, 247) = 3.40$, $\eta^2 = .04$, $p < .05$). Independent-samples t-tests revealed that African American participants had higher game ratings (mean difference = 1.13, $p < .05$), and marginally higher progress ratings (mean difference = 0.68, $p < .1$) in the role model condition. White participants reported lower engagement in the role model condition (mean difference = 0.45, $p < .01$). South Asian players reported higher engagement in the role model condition (mean difference = 0.55, $p < .05$). See Figure 5 for illustration.

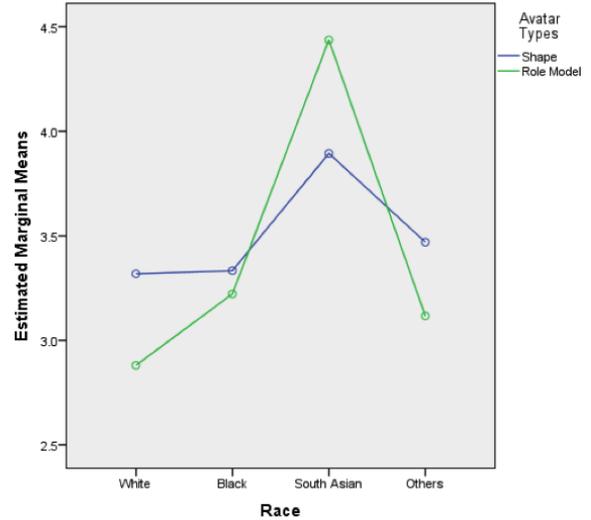


Figure 5: Average Enjoyment.

Table 1: Players completing ≤ 1 levels.

Attribute	R-Mean	R-SD	S-Mean	S-SD	t-test
Levels Completed	0.52	0.50	0.44	0.50	1.05
Total Bonus Items	0.83	1.33	0.39	1.00	2.56*
Average Enjoyment	3.23	1.34	3.36	1.11	0.49
Avatar Rating	4.11	0.86	3.55	0.95	4.28**

* $< .05$, ** $< .01$, R = Role Model, S = Shape, SD = Standard Deviation

Table 2: Participants selecting same gender role models versus different gender role models.

Attribute	Rs-Mean	Rs-SD	Rd-Mean	Rd-SD	t-test
Levels Completed	1.43	1.07	1.23	1.11	1.00
Total Bonus Items	2.23	2.90	1.67	2.55	1.10
Average Enjoyment	3.43	1.20	3.32	1.18	0.43
Avatar Rating	4.21	0.84	4.08	0.77	0.91

* $< .05$, ** $< .01$, Rs = Same Gend., Rd = Diff. Gend., SD = Standard Deviation

Performance Split

Players that completed few levels collected more bonus items in the role model condition. Players that completed ≤ 1 levels collected more bonus items in the role model condition, $p < .05$, $r = 0.19$. See Table 1.

Same vs. Different Gender Role Models

No significant differences were found. See Table 2.

Gender Selections: A chi-square test was used to determine whether there was a significant difference between male and female participants and the chosen role model's gender. 76.5% of female role models were chosen by female participants. 79.1% of male role models were chosen by male participants. The difference between male and female participants was significant, $\chi^2 = 39.63$, $df = 1$, $p < .001$.

Table 3: Participants selecting same race role models versus different race role models.

Attribute	<i>Rs</i> -Mean	<i>Rs</i> -SD	<i>Rd</i> -Mean	<i>Rd</i> -SD	t-test
Levels Completed	1.38	1.11	1.39	1.03	0.03
Total Bonus Items	2.27	3.06	1.84	2.39	1.00
Average Enjoyment	3.47	1.21	3.32	1.17	0.70
Avatar Rating	4.31	0.79	3.97	0.83	2.73**

* $<.05$, ** $<.01$, *Rs* = Same Race, *Rd* = Diff. Race, SD = Standard Deviation

Same vs. Different Race Role Models

Players had higher avatar ratings for same race role models. Players reported a higher avatar rating for same race role models, $p < .01$, $r = 0.20$. See Table 3.

Race Selections: A chi-square test was used to determine whether there was a significant difference in the chosen role model’s race. Participants tended to pick a similar race role model. The difference was statistically significant, $\chi^2 = 115.52$, $df = 9$, $p < .001$.

Between Role Model Types

A cross tabulation was checked for any difference between participants’ performance across role model types. The Chi-square test indicated no significant difference ($\chi^2 = 4.29$, $df = 11$, $p > .05$) between high and low performing participants across the 12 different role model types (see Table 4).

Discussion

The results suggest that role model avatars can enhance performance and engagement for some groups of participants. African American participants had higher game affect in the role model condition. South Asian participants had higher engagement in the role model condition. Participants completing ≤ 1 levels had higher performance in the role model condition. Therefore, an AI system that generates avatars would do well to utilize both the player demographics and the avatar type.

Participants in the user role model condition rated same race role models as higher. This, and the general trend observed in Tables 2 and 3, supports the literature, i.e., role models of similar gender and race appear to be more effective (Marx and Goff 2005; Blanton, Crocker, and Miller 2000). Unexpectedly, we found that white participants had lower reported engagement in the role model condition. Previous studies have reported that role models can improve the academic performance of some social groups (i.e., female and African American participants), so while we would expect little to no effect in white participants, we see the *opposite* effect. While our current analyses cannot conclude as to why this occurred, this is an indication that role models may not *always* be effective. For instance, one possible alternative explanation is cultural differences (e.g., a human photo may appear out of place). In addition, “superstar” role models can cause self-deflation (Lockwood and Kunda 1997).

Indeed, this is a complex topic; if we had attempted to link effects to specific types of role models (e.g., scientists, athletes, etc.), it is unclear whether the effect is due to the

Table 4: Performance by role model types

Role Model Type:	High/Low performing groups		Total
	Low Performing	High Performing	
Actor	61.1%	38.9%	100.0%
Scientist	37.5%	62.5%	100.0%
TV personality	50.0%	50.0%	100.0%
Astronaut/Pilot	50.0%	50.0%	100.0%
Athlete	48.1%	51.9%	100.0%
Author	50.0%	50.0%	100.0%
Fictional Character	56.4%	43.6%	100.0%
Magnate	40.0%	60.0%	100.0%
Musician/Singer	61.9%	38.1%	100.0%
Political Figure	58.8%	41.2%	100.0%
Religious Figure	75.0%	25.0%	100.0%
Other	44.4%	55.6%	100.0%

type of role model, the type of person that picks that type of role model, or both. More targeted studies are needed to explore specific role models. Our results expand upon the findings of the social science literature to also include role model *avatar* as a possible means of enhancing player engagement and performance.

Conclusion

This paper has presented a comparative exploratory study between player role model avatars and player geometric shape avatars. Player role model avatars resulted in higher self-reported engagement and performance compared to player shape avatars in several categories of players. Educational games could be pre-populated by a set of role model avatars to better support players’ engagement and performance. These results help provide the basis of a follow-up project to develop personalization algorithms for adaptive learning systems that dynamically adapt the virtual identities of students to support performance, engagement, and learning within a broader ecology.

Future Work

We have pending funding to bring these studies into public schools. Also, funding is in place to study these effects in an international setting. We are currently investigating role model professions. For example, an experiment with conditions corresponding to role model types. Also, blending aspects of the user with aspects of a role model could potentially be effective, e.g., an avatar wearing a t-shirt of the role model or that is assigned some type of tags or points associated with the role model or the role model’s profession.

Acknowledgments

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

- Aronson, J.; Fried, C. B.; and Good, C. 2002. Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*.
- Blanton, H.; Crocker, J.; and Miller, D. T. 2000. The Effects of In-Group versus Out-Group Social Comparison on Self-Esteem in the Context of a Negative Stereotype. *Journal of Experimental Social Psychology* 36(5):519–530.
- Block, M., and Siegel, R. 2009. Duncan: 'Educate Our Way To A Better Economy'.
- Buunk, A. P.; Peiró, J. M.; and Griffioen, C. 2007. A positive role model may stimulate career-oriented behavior. *Journal of Applied Social Psychology* 37:1489–1500.
- Cox, T. H., and Blake, S. 1991. Managing cultural diversity: implications for organizational competitiveness.
- Dominguez, I. X., and Roberts, D. L. 2015. Asymmetric Virtual Environments : Exploring the Effects of Avatar Colors on Performance. *Experimental Artificial Intelligence in Games: Papers from the AIIDE Workshop*.
- Gee, J. P. 2007. *What Video Games Have to Teach Us About Learning and Literacy*. Palgrave Macmillan.
- Genskow, L. R. 2006. Challenges and Opportunities in Process Innovation.
- Good, C.; Rattan, A.; and Dweck, C. S. 2012. Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of personality and social psychology*.
- Harrell, D. F.; Kao, D.; Lim, C.; Lipshin, J.; and Sutherland, A. 2014. The Chimeria Platform: User Empowerment through Expressing Social Group Membership Phenomena. *Digital Humanities*.
- Harrell, D. F. 2010. Toward a Theory of Critical Computing: The Case of Social Identity Representation in Digital Media Applications. *CTheory*.
- Harrell, D. F. 2013. *Phantasmal Media: An Approach to Imagination, Computation, and Expression*. The MIT Press.
- Kao, D., and Harrell, D. F. 2015a. Exploring construction, play, use of virtual identities in STEM learning. *Jean Piaget Society Annual Conference*.
- Kao, D., and Harrell, D. F. 2015b. Mazzy: A STEM Learning Game. *Foundations of Digital Games*.
- Kao, D., and Harrell, D. F. 2015c. Toward Avatar Models to Enhance Performance and Engagement in Educational Games. *Computational Intelligence in Games*.
- Kao, D., and Harrell, D. F. 2015d. Toward Evaluating the Impacts of Virtual Identities on STEM Learning. *Foundations of Digital Games*.
- Lim, C.-U., and Harrell, D. F. 2013. Modeling Player Preferences in Avatar Customization Using Social Network Data. In *Computational Intelligence in Games*.
- Lim, C., and Harrell, D. 2014. Developing Social Identity Models of Players from Game Telemetry Data. *AIIDE*.
- Linebarger, J. M., and Kessler, G. D. 2002. The effect of avatar connectedness on task performance. *Lehigh Univ TR*.
- Lockwood, P., and Kunda, Z. 1997. Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology* 73(1):91–103.
- Lockwood, P. 2006. "Someone like me can be successful": Do college students need same-gender role models? *Psychology of Women Quarterly* 30(1):36–46.
- Marx, D. M., and Goff, P. A. 2005. 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* 44:645–657.
- Marx, D. M., and Roman, J. S. 2002. Female Role Models: Protecting Women's Math Test Performance. *Personality and Social Psychology Bulletin* 28:1183–1193.
- Marx, D. M.; Stapel, D. a.; and Muller, D. 2005. We can do it: the interplay of construal orientation and social comparisons under threat. *Journal of personality and social psychology* 88(3):432–446.
- McIntyre, R. B.; Lord, C. G.; Gresky, D. M.; Eyck, L. L. T.; and Bond, C. F. 2005. A Social Impact Trend in the Effects of Role Models on Alleviating Women's Mathematics Stereotype Threat. *Current Research in Social Psychology* 10(9):1–26.
- McIntyre, R. B.; Paulson, R. M.; and Lord, C. G. 2003. Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology* 39(1):83–90.
- Merton, R. K. 1936. The Unanticipated Consequences of Purposive Social Action.
- Morie, J. F., and Verhulsdonck, G. 2008. Body/Persona/Action! Emerging Non-anthropomorphic Communication and Interaction in Virtual Worlds. *Advances on Computer Entertainment Technology* 365.
- Papert, S., and Harel, I. 1991. Situating Constructionism.
- Papert, S. 1993. *Mindstorms*.
- Perlin, K. 2002. <http://mrl.nyu.edu/~perlin/polly/track.html>.
- Piaget, J. 1976. *Piaget and His School*. Springer Berlin Heidelberg.
- Shih, M.; Pittinsky, T. L.; and Ambady, N. 1999. Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance.
- Steele, C., and Aronson, J. 1995. Stereotype Threat and the Intellectual Test Performance of African Americans. *Journal of personality and social psychology*.
- Steele, C. 1997. A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist* 52(6):613–629.
- Stricker, L. J., and Ward, W. C. 2004. Stereotype Threat, inquiring About Test Takers Ethnicity and Gender, and Standardized Test Performance. *Journal of Applied Social Psychology* 34(4):665–693.
- Waytz, A.; Heafner, J.; and Epley, N. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52:113–117.
- Yee, N., and Bailenson, J. 2007. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Comm. Research*.
- Yee, N.; Ducheneaut, N.; Yao, M.; and Nelson, L. 2011. Do Men Heal More When in Drag? *CHI 2011* 1–4.