

University of Memphis

University of Memphis Digital Commons

---

Electronic Theses and Dissertations

---

7-16-2012

## Users' Preference for and Perception of Animated Pedagogical Agents

Kyle Randall Cheney

Follow this and additional works at: <https://digitalcommons.memphis.edu/etd>

---

### Recommended Citation

Cheney, Kyle Randall, "Users' Preference for and Perception of Animated Pedagogical Agents" (2012). *Electronic Theses and Dissertations*. 533.  
<https://digitalcommons.memphis.edu/etd/533>

This Thesis is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact [khhgerty@memphis.edu](mailto:khhgerty@memphis.edu).

USERS' PREFERENCE FOR AND PERCEPTION OF ANIMATED PEDAGOGICAL  
AGENTS

by

Kyle Randall Cheney

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Psychology

The University of Memphis

August 2012

## **Abstract**

Cheney, Kyle Randall. MSGP. The University of Memphis. August/2012. Users' Perceptions of and Preference for Animated Pedagogical Agents. Major Professor: Arthur C. Graesser.

This thesis investigates the effect that ethnicity and gender of animated pedagogical agents (APAs) has on preference and perception by participants. It was hypothesized that participants would choose to work with and give more favorable ratings to those agents who most resemble themselves. Ratings of four professor agents, an African American male and female and a Caucasian male and female, were collected from 120 workers on Mechanical Turk, an online crowd sourcing marketplace. Ethnicity and gender of the agents were externally validated before use in the study. Results indicated that users did not prefer agents who were identical to themselves. Instead, the results followed trends of previous research in the stereotyping research on students' ratings of professors. Results indicated that stereotypes for African American and Caucasian professors may be more of a guiding factor in participants' feelings towards APAs.

## TABLE OF CONTENTS

Section		Page
Introduction		1
Methodology		5
Results		9
Discussion		13
References		18
Appendices		
A	Example Pay Scale for Workers	23
B	Screen Shot of Mechanical Turk Tasks	24
C	Informed Consent	25
D	Script used in the Study	26
E	Gender and Ethnicity Validation Measure	27
F	Demographic Information Survey	28
G	Agent Persona Inventory	29
H	Measure of Affability	30
I	Images of the Agents for the Study	31
Tables		32

## List of Tables

Table		Page
1	The Average Score for all Four Agents on Each of the Four Validation Measures.	32
2	Pooled and Weighted Values and Significance Levels for the Four Similar and 12 Dissimilar Means.	33
3	The Mean Score for the Measure of Human-Like.	34
4	The Mean Score for the Measure of Engaging.	35
5	The Mean Score for the Measure of Credibility.	36
6	The Mean Score for the Measure of Affability.	37
7	The Mean Score for the Measure of Facilitates Learning.	38
8	The Results the Chi Square Test of Agent Choice by Participants.	39

## **Users' Preference for and Perception of Animated Pedagogical Agents**

We have all seen Clippy, the talking paper clip of Windows, or a helpful in-game character who can guide us to the next level in a video game. What makes these characters effective at their tasks (or not)? What makes us prefer one character over another? In our quest for more effective and more cost efficient educational materials, we have come to rely heavily on technology. One feature that many tutoring systems and other advanced learning environments have in common is an on-screen character to help facilitate knowledge acquisition, called an animated pedagogical agent (APAs). APAs have been defined in various ways. In many definitions, there is a prototypical set of characteristics that can be succinctly stated: an APAs is an animated, life-like agent designed to facilitate learning in a multimedia environment (Lester et al., 1997; Moreno, Mayer, Spires, & Lester, 2001; Wouters, Pass, & van Merriënboer, 2008). APAs are one way in which education has incorporated technology in an attempt to enrich the experience and create the most effective learning technologies possible. For this reason research focusing specifically on the creation and implementation of these tools is important.

Examples of these systems are AutoTutor (Graesser, Jeon, & Dufty, 2008), Operation ARIES (Millis et al., 2011), Guru (Olney et al., 2012), Betty's Brain (Biswas, Jeong, Kinnbrew, Sulcer, & Roscoe, 2010), Steve (Johnson, Rickel, & Lester, 2000), and Herman the Bug (Lester, Stone, Stelling, 1999). Many of these are essentially talking, animated heads with facial expressions and artificial speech generation. These characters are much more advanced versions of the assistance based characters mentioned previously (e.g., Clippy). These characters are more dynamic and can be viewed as companions during the learning process.

In an ever advancing technological world, understanding the nuances of the educational multimedia we are exposed to is crucial. But with the increasing presence of technology in our everyday lives, if educational product designers hope to capture the attention of the users when using APAs, they will need to understand what aspects the user may be most attracted to and what effects they have on users.

### **Previous Research on Human-Computer Interaction**

Nass and colleagues have repeatedly shown that humans treat computers as social entities that operate under the same or similar rules as human-human interactions (Ibster & Nass, 2000; Lee & Nass, 1998; Nass, Steuer, & Tueber, 1994; Reeves & Nass, 1996). In fact, research has shown that with even basic cues or characteristics, such as giving the computer a voice, participants will apply gender stereotypes to a computer based on that voice (Nass, Moon, & Green, 1997).

One significant finding is that computers can be counted as members of arbitrary groups. Nass and Moon (2000) found that by having participants wear armbands of the same color as that marked on a computer (as opposed to an alternative color), the level of cooperation and conformity to the computer's suggestions increased. This suggests that participants included the computer as part of their arbitrary group. It became a member of the group just by it having the same color markings. This study showed that people naturally impose social interaction rules on computers even without any major anthropomorphized features to cue such rules.

### **Past Research on Animated Pedagogical Agents**

Research on APAs has produced mixed results with some researchers questioning the benefits of using APAs at all. One such argument is that the richness of the agent may

create a cognitive overload situation (Clark & Choi 2007). Cognitive overload refers to a situation in which the amount of information being processed simultaneously exceeds the amount of available working memory resulting in the loss of information. However, research has provided compelling evidence that this is not always the case. Mayer (2001), for example, found evidence that when presenting information using pictures or other multimedia channels, the words should be spoken, rather than written, in order to minimize interference with the processing of the images. This is compatible with theories of dual-coding in processing and storing information (Paivio, 1986) which state we process auditory and visual information through different cognitive channels. Although Mayer's research was conducted using a voice over and not agents, other work on this subject has supported claims that we process information through multiple channels and that having an agent is beneficial to learning (Craig, Gholson, & Driscoll, 2002; Moreno et al., 2001; Pavio, 1986; Wouters et al., 2008).

One major flaw in the research on APAs is the way the agents are designed with respect to physical characteristics. Studies on the effects of the ethnicity and gender of the agent often involve ethnicity and gender being decided by the designer with no external validation. That is, only one study has reported using external validation to assure the ethnicity and gender of the agents were perceived as what the researchers' intended (Baylor & Kim, 2003). Another flaw is representativeness of sample. In one study on users' perception of agents, the sample was 90% Caucasian (Pratt, Hauser, Ugray, & Patterson, 2007) which fails to accurately represent the greater population.

Many of the studies used characters that were generated with and powered by software which may have been current at the time, but is now outdated, such as Microsoft



Agent (Microsoft, 2), Poser (Smith Micro Software, 5, although newer versions of Poser are now available), or Mimic pro2 (2.1). As technologies progress, researchers should continue to investigate their effects and also ensure that the results of previous studies are not merely artifacts of problematic software. In a recent study on agent quality (Cheney, Germany, Fike, Craig, & Gholson, 2009), agents were created using software such as People Putty (Hapttek, 1) and Authorware (Adobe, 7.2) and compared to those created by Microsoft Agent. The agents were of the same gender and ethnicity in both conditions. The results showed that the newer agents were perceived as significantly more lifelike, credible, and engaging. The participants who interacted with the newer agents also showed significantly higher learning gains on both deep and shallow problems. Given the advancement in technology and its significant effect in perception of APAs, it is important to have an evaluation of previous claims to ensure the results are not artifacts of the agents themselves and that the social implications of ethnicity and gender are still relevant.

### **The Current Study**

The goal of this thesis was to investigate participants' perception of and preferences for APAs by surveying a demographically diverse group of people about APAs that were externally validated on both gender and ethnicity. This thesis served as a replication of previous research and a test of those theoretical claims generated in psychology and communication. More specifically, the two hypotheses below were tested in the thesis.

**Hypothesis H1: When given the choice, users will choose agents who are most like themselves with regard to ethnicity and gender.** Research on choice of agents has

shown that non-white users are more likely to choose agents that are similar to themselves compared with white users (Baylor, 2005; Moreno & Flowerday, 2005). This research is further supported by the Similarity-Attraction hypothesis (Byrne, Griffitt, & Stefaniak, 1967) that states we are attracted to those people who are most similar to ourselves.

**Hypothesis H2: Participants will give higher ratings to agents who are most similar to themselves with regard to ethnicity and gender.** Baylor and Kim (2003) investigated the effect of gender and ethnicity on users' perceptions of APAs. Their results indicated that users rate agents of the same ethnicity as more engaging and affable. This research is also supported by the Similarity-Attraction hypothesis.

## **Methods**

### **Participants**

All participants were collected using Mechanical Turk, an online crowd sourcing marketplace. Requesters (researchers) are able to post tasks online for workers (participants) to complete. The workers are anyone over the age of 18 from all over the world. However, for this study the workers were restricted to those from the United States.

When tasks are completed, the requester pays the worker for their time. A set price is posted for completion of each task. Pay is directly proportional to the time the task takes to complete and is at the discretion of the requester. For this study, workers were paid at a rate of minimum wage (\$7.25/hour). An example of the pay scale is shown in Appendix A. A screen capture of the research tasks in Mechanical Turk is shown in Appendix B.

**Informed Consent.** Because Mechanical Turk allows for qualifications to be set, informed consent (Appendix C) was collected by the informed consent being posted as a qualification for completing the other tasks. Completion of the task (selecting “I agree to participate in the study”) served as acknowledgement and consent. Only after completing this task were the participants allowed access to the study.

**Demographics.** Data for 120 participants were collected but seven declined to answer information about their ethnicity and/or gender and were therefore unable to be included in the main analysis. The ethnicity and gender breakdown for the remaining 113 participants was: 24 African American Males, 33 African American Females, 30 Caucasian Males, and 36 Caucasian Females. The age of the workers ranged from 18 to 63 years ( $M = 31$ ,  $SD = 10.17$ ).

## **Design**

This study used a factorial design with four independent variables: 2 (student gender) x 2 (student ethnicity) x 2 (agent gender) x 2 (agent ethnicity). Agent ethnicity and gender were repeated measures variables whereas student ethnicity and gender were between-subjects variables.

## **Materials and Counterbalancing**

**Agents.** The agents were designed with the program Character Builder (Media Semantics, 5.3.3). The text-to-speech engines that provided the voices for the agents were Paul (NeoSpeech, 1) and Kate (NeoSpeech, 1). The way the agents were presented was via a short video in which the agents introduced themselves and explained that they would be the participants’ professor. The videos were identical with regards to what the agent said and all facial expressions. The script and facial expressions is provided in

Appendix D. These videos were rendered directly from Character Builder into an MP4 file. They were embedded into the task posted on Mechanical Turk. There was one African American male, one African American female, one Caucasian male, and one Caucasian female agent. The selection of agents is described in the Measures section below.

**Counter balancing and randomization.** In order to eliminate any ordering effects that may occur for an agent appearing before or after any other agent, a Latin square design was used for counterbalancing. Each participant received the agents (numbered 1, 2, 3, or 4) in one of four orders (i.e., 1234, 2143, 2413, & 4321). Each order had a specific number of participants to ensure that each order was exposed to an equal number of participants (30 for each order). Participants were randomly assigned to one of the four orders.

### **Measures**

***Agent Validation.*** Previous validation measures on gender and ethnicity were conducted by comparing ratings on these dimensions. For example, an agent was considered male if it was rated significantly more male than female agents (Baylor & Kim, 2003). In order to ensure the agents were being rated on their own characteristics and not in comparison to other agents, validation of these categories was conducted with normative study. Each agent was rated on levels of (a) femininity and masculinity and (b) African American-ness/Caucasian-ness. If, for example, an agent was intended to be female, it would still be rated on both its level of masculinity and femininity because an agent could be perceived as a hybrid of the two genders or be ambiguous. The measures of gender and ethnicity were based on a 6-point anchored scale as shown in Appendix E.

In order to make the requirements for achieving the desired attribute slightly more stringent, rather than using a simple midpoint (i.e., 3.5 for this scale) cut-off, agents had to achieve mean scores above 4 on desired variables and below 3 on non-desired variables. For instance, in order for the agent to be considered female, it had to receive a mean score of less than 3 on the femininity dimension and greater than 4 on the masculinity dimension. The same process was used for ethnicity. Each of the four agents was rated by at least 40 workers on Mechanical Turk.

***Demographic information.*** After completion of each task, the participants were asked to complete a basic demographic survey that is provided in Appendix F.

***Rating of agents' persona.*** An agent's persona was measured using the Agent Persona Inventory (API) (Ryu & Baylor, 2005). The inventory has four subscales: facilitating learning (10 questions), humanlike (5 questions), credible (5 questions), and engaging (5 questions). The original API has a Likert scale response system for each question on a 1 to 5 scale. To avoid users answering neutrally, a 1 (Strongly Disagree) to 6 (Strongly Agree) scale was adopted in this study that anchors the numbers and segregates agree versus disagree ratings. The list of the questions on that survey is given in Appendix G.

***Rating of agents' level of affability.*** A three-item survey (Baylor & Ryu, 2003) was administered on the agents' levels affability. It asked the users to rate the agent on niceness, attractiveness, and how personally warm the agent is (Appendix H). To avoid users answering neutrally, a 1 (Strongly Disagree) to 6 (Strongly Agree) anchored scale was used.

*Agent preference.* All four agents were presented to participants in a row. The participants were asked to choose the one agent with who they would like to be their professor on a learning task (Appendix I).

## **Procedure**

The first step was to post a qualification ranking in Mechanical Turk to implement informed consent. Once participants completed the qualification ranking, which informed them of their rights and the intent to use the anonymous data collected for research, they were able to access the other tasks.

Once the agents were validated selected via the process described above, collection of the data for analysis began. There were two types of tasks, one for evaluating each of the hypotheses. For the independent rating tasks, the users viewed a short video, approximately 30 seconds long, in which the agents introduced themselves. Below the video were the measures of agent persona and affability. The agents were introduced and rated, one agent at a time, following the counterbalancing scheme.

Following the rating task was the agent preference task. All the agents were presented side-by-side in the same order as they were presented for the independent rating tasks. The participants were asked to choose which agent they would like to be their professor.

Finally, the user provided demographic information as shown in Appendix F.

## **Results**

### **Agent Validation**

Each agent was submitted to the validation procedure. Table 1 displays the mean score for each agent on each of the variables. Since no agent was outside the determined

parameters (above 4 on desired variables and below 3 on non-desired variables), there was no need to create more agents. The agents used in the study are presented in Appendix I.

### **Agent Perception**

To test the hypothesis that participants would rate agents who are more similar to themselves higher than agents who are more dissimilar, a contrast was conducted comparing the means of all four similar-to-participant ratings with the means of the twelve dissimilar-to-participant ratings for each of the five dependent variables (the four factor loadings of the API (Facilitates Learning, Credible, Human-Like, and Engaging) and the measures of affability) by pooling the means. Not only was the comparison was not significant on any of the measures, but on four out of five of the measures the means were in the opposite direction than was hypothesized. The means and  $F$  statistics are reported in Table 2.

In order to investigate the presence of any interactions that may be present among the data, mixed ANOVAs were conducted with agent ethnicity and gender as the repeated measures and the participants' ethnicity and gender as the between-subjects variables. This was completed for each of the four API factor loadings (Human-Like, Engaging, Credible, and Facilitates Learning) and for the measures of affability. The findings for each are discussed below.

***Human-Like.*** There were no significant main effects or interactions among the variables on the measures of how human-like the agents were. This outcome is promising because it indicates that no agent was significantly more realistic than any other agent. Given the persona effect (Lester et al., 1997), which states that users find more life-like

agents more helpful, credible, and entertaining, this lends a level of control and validity to any significant findings that emerge with the other variables. Table 3 displays the overall means for this variable.

**Engaging.** Agents' ethnicity had a significant effect on how engaging the agent was rated by participants,  $F(1, 109) = 10.90, p = .001, \eta^2 = .09$ . Specifically, Caucasian agents ( $M = 3.65, SD = .87$ ) were rated as more engaging than were African-Agents ( $M = 3.50, SD = .89$ ).

There was also a significant interaction between the agents' ethnicity, the ethnicity of the participants, and the gender of the participants,  $F(1, 109) = 5.26, p = .024, \eta^2 = .04$ . A series of post hoc, independent sample  $t$ -tests revealed the interaction to be between ratings of the Caucasian agents by African American females and Caucasian males,  $t(61) = .033, p = .003, d = .55$ . Specifically, African American females ( $M = 3.87, SD = .88$ ) rated the Caucasian agents higher than did the Caucasian males ( $M = 3.41, SD = .79$ ). Table 4 shows the overall means for this variable.

**Credibility.** Agents' ethnicity had a significant effect on how credible the agent was rated by participants,  $F(1, 109) = 4.82, p = .03, \eta^2 = .04$ . For credibility, it was the case that the Caucasian agents ( $M = 3.18, SD = .08$ ) were rated higher than African American agents ( $M = 3.05, SD = .08$ ) by participants. There was also an interaction between the agents' ethnicity and the ethnicity of the participants,  $F(1, 109) = 5.81, p = .018, \eta^2 = .05$ . Results of a post hoc, paired sample  $t$ -test ( $t(56) = 3.29, p = .002, d = .37$ ) indicated that African American participants rated the Caucasian agents ( $M = 3.29, SD = .76$ ) significantly more credible than the African American agents ( $M = 2.99, SD = .83$ ). There was no difference ( $t(55) = .32, p = .74$ ) in how Caucasian



participants rated Caucasian agents ( $M = 3.14$ ,  $SD = 1.00$ ) and African American agents ( $M = 3.11$ ,  $SD = 1.05$ ). Table 5 shows the overall means for this variable.

**Affability.** On the measure of affability, there were no main effects but there were two interactions. The first interaction was between agent ethnicity and the ethnicity of the participants,  $F(1, 109) = 5.83$ ,  $p = .017$ ,  $\eta^2 = .05$ . Post hoc, paired sample  $t$ -tests ( $t(56) = -2.51$ ,  $p = .015$ ,  $d = .51$ ) indicated that African American participants rated the Caucasian agents ( $M = 3.22$ ,  $SD = .85$ ) significantly more affable than the African American agents ( $M = 2.78$ ,  $SD = .86$ ) while the Caucasian participants showed no significant difference ( $t(55) = .50$ ,  $p = .50$ ) in their ratings of Caucasian agents ( $M = 3.03$ ,  $SD = .84$ ) and African American agents ( $M = 3.10$ ,  $SD = .87$ ).

The second interaction was between agents' gender and the gender of the participants,  $F(1, 109) = 5.78$ ,  $p = .018$ ,  $\eta^2 = .05$ . Post hoc, paired sample  $t$ -tests ( $t(56) = 2.48$ ,  $p = .016$ ,  $d = .33$ ) indicated that male participants rated the male agents ( $M = 3.24$ ,  $SD = .86$ ) significantly more affable than the female agents ( $M = 2.96$ ,  $SD = .80$ ). There was no significant difference in how female agents rated the male and female agents on affability. Table 6 shows the overall means for this variable.

**Facilitates Learning.** There were no main effects on the measure of the agents' ability to facilitate learning, but there were two significant interactions. First, there was an interaction between the agents' ethnicity, the ethnicity of the participants, and the gender of the participants,  $F(1, 109) = 4.04$ ,  $p = .047$ ,  $\eta^2 = .03$ . A series of post hoc, independent sample  $t$ -tests revealed the interaction to be between ratings of the Caucasian agents by African American females and Caucasian males,  $t(61) = 2.22$ ,  $p = .03$ ,  $d = .57$ . Specifically, African American females ( $M = 3.84$ ,  $SD = .86$ ) rated the Caucasian agents

as better able to facilitate learning than did the Caucasian males ( $M = 3.32$ ,  $SD = .96$ ). No other participant group (African American females, Caucasian males, or Caucasian females) showed differences in how they rated Caucasian and African American Agents.

Second, there was an interaction between the agents' ethnicity, the agents' gender, and the gender of the participants,  $F(1, 109) = 5.03$ ,  $p = .027$ ,  $\eta^2 = .04$ . A post hoc repeated measures ANOVA revealed the interaction to be that African American participants rated the Caucasian female agent ( $M = 3.70$ ,  $SD = .95$ ) significantly higher than all other agents (Caucasian male:  $M = 3.66$ ,  $SD = 1.01$ ; African American male:  $M = 3.44$ ,  $SD = 1.06$ ; African American female:  $M = 3.45$ ,  $SD = 1.11$ ). There was no significant difference in Caucasian students' ratings of how able to facilitate learning each agent was. Table 7 shows the overall means for this variable.

### **Agent Preference**

Non-parametric tests were used to investigate any significant differences in preference of agents. For both female participant groups the difference in choice was not significant (Caucasian Females:  $\chi^2(3, N = 26) = .76$ ,  $p = .857$ ; African American females:  $\chi^2(3, N = 33) = 1.78$ ,  $p = .618$ ). However, the difference in choice was significant for both male participant groups (Caucasian males:  $(\chi^2(3, N = 30) = 13.20$ ,  $p = .004$ ; African American males:  $(\chi^2(3, N = 24) = 9.00$ ,  $p = .029$ ). For both of the male groups, the preferred agent was the Caucasian Female and not the agent who was most similar to themselves. Table 8 reports the results of this test.

### **Discussion**

The findings of this study indicate that there is a much more complicated relationship between the ethnicity and gender of participants and their ratings of APAs

than a simple similarity hypothesis. Planned comparison tests indicated that the similarity hypothesis was not sufficient enough to predict the preference for and perception of APAss. The results indicated that users did not rate the agents who were most like themselves significantly higher than agents who were most dissimilar to themselves on any of the measures. In fact, the means of the comparison were in the opposite direction on four out of five means (Table 2). This is further complicated by the seemingly contradictory information provided by the choice task. While women seemed not to have a preference for which agent they would like to work, both Caucasian and African American men overwhelmingly chose to work with Caucasian female agent. However, on almost all rating measures the Caucasian men rated that agent the lowest.

A deeper look into the results indicated there were significant interactions between the ethnicity and gender of the participant and the ethnicity and gender of the agents. Gender of the agent also appears to be a less powerful indicator than was ethnicity. Specifically, the ethnicity of the agent was involved in the interactions on all measures that were significant, while the gender of the agent was only involved on measures of affability and ability to facilitate learning. In contrast, both ethnicity and gender of the participant were involved in interactions on all measures except credibility, where only the students' ethnicity was involved.

These results offer little support for the hypotheses that users prefer APAs that are more similar to themselves and will rate them more highly than agents who are more dissimilar than themselves. Interestingly, a common trend in the interactions was that the African American participants rated the Caucasian agents more highly than the African American agents. These findings contradict findings that non-white users are more likely

to choose to work with agents who are most similar to themselves (Moreno & Flowerday, 2005). While this is in direct conflict with the hypotheses and previous research on agents, research into stereotypes held for professors and the evaluations that students provide for professors may explain the findings, as discussed below.

In this study, Caucasian participants did not show differences in their ratings of Caucasian and African American agents. However, these results fit with previous research in which Caucasians judged other Caucasians more harshly than member of other groups. This is known as the expectancy effects or a desire to appear accepting (Bettencourt, Dill, Greathouse, Charlton, & Mulholland, 1997; Jones et al., 1984). Taken together, this suggests that the ratings by Caucasian participants may experience a sort of regression towards the mean effect, drawing the rating of both African American and Caucasian agents towards the middle.

Research on professors' evaluations is helpful in suggesting why the African American participants consistently rated Caucasian agents higher than African American agents. In a study using observation, survey, and interviews as the methods, Hendrix (1998) found that both Caucasian and African American students used different criteria to evaluate African American professors than they did for Caucasian professors. Specifically, the criteria were much more stringent for the African American professors, where subject matter and ethnicity (rather than ethnicity alone) interacted to inform the evaluation, resulting in lower scores. Similarly, in this study, when the difference in rating was significant, it was the Caucasian agents who were more highly.

A recent analysis of the student-generated evaluations on the site RateMyProfessor.com (Reid, 2010) reported that African American and Asian professors

were significantly more likely to receive lower ratings than were the white professors. Further, the African American male professors were rated the lowest. This is consistent with the findings of this study.

Together these findings on professor evaluations and stereotyping indicate that it is much more likely that the perception of and preference for APAs is guided by stereotypes held by users rather than their own personal characteristics. While one study did find that Caucasians' initial prejudices can be overcome through extensive interaction and the building of a relationship (Jackson & Crawley, 2003), this is somewhat impractical in the world of agent-based learning systems as most of them are designed for short interactions, though hopeful for traditional classroom settings. For this reason, it is important to further examine the effects of these stereotypes on APAs and the preferences of students in selecting agents. The limitations of this study should inform future investigations into this topic. Specifically, with the knowledge that stereotypes are so powerful, gaining insight into the stereotypes held by participants seems prudent. Adding a survey on the beliefs of the participants may offer further insight into the way stereotypes effect the decisions of participants.

Also, while this study was designed to specifically investigate the relationship of Caucasians and African Americans and had a diverse sample of those populations, it, like many studies on minorities, left out Hispanics, which is, as shown in of the 2010 U.S. Census, the largest minority group in the United States with 16.3% of the population. Expanding the ethnicity of both the agents and the participants to include this group, and perhaps others, may reveal other interactions.

Further, there may be other factors that affect the ratings and choice of agents. In the stereotype literature, there is an interaction between ethnicity and gender with regard to women and minorities. Research has shown that for female professors, but not males, to be rated as competent they have to also show characteristics of warmth and have significant social contact with the students (Kierstead, D'Agostino, & Dill, 1988).

Many of these effects may be best studied on full scale teaching environments. Given the short interaction with the agents, the measures in this study were based solely on first impressions and stereotypes. If given the chance to interact more with the agents, perhaps the stereotypes would play a less significant role or would be mitigated by the subject matter, attitude, or other features that can only be manipulated fully in longer interactions.

## References

- Authorware (Version 7.2) [Computer Software]. San Jose, CA: Adobe
- Baylor, A., & Kim, Y. (2003). The role of gender and ethnicity in pedagogical agent perception. In G. Richards (Ed.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 1503-1506). Chesapeake, VA: AACE.
- Bettencourt, B. A., Dill, K. E., Greathouse, S. A., Charlton, K., & Mulholland, A. (1997). Evaluations of ingroup and outgroup members: The role of category-based expectancy violation. *Journal of Experimental Social Psychology, 33*, 244 - 275.
- Biswas, G., Jeong, H., Kinnebrew, J., Sulcer, B., & Roscoe, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology-Enhanced Learning, 5*(2), 123-152.
- Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology, 5*, 82-90.
- Character Builder (Version 5.3.3) [Computer Software]. Snoqualmie, WA: Media Semantics
- Cheney, K. R., Germany, M., Fike, K., Craig, S. D., & Gholson, B. (2009). *The role of agent quality in vicarious learning*. Presented at the 2009 Annual APS convention in San Francisco.
- Clark, R., & Choi, S. (2007). The questionable benefits of pedagogical agents: Response to Veletsianos. *Journal of Educational Computing Research, 36* (4), 379-381.

- Craig, S. D., Gholson, B., & Driscoll, D. (2002.) Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*, 428-434.
- Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes, 45*, 298-322.
- Hendrix, K. G. (1998). Student perceptions of the influence of race on professor credibility. *Journal of Black Studies, 28*, 738 –763
- Isbister, K., & Nass, C., 2000. Consistency of personality in interactive character: Verbal cues, non-verbal cues, and user characteristics. *International Journal Human–Computer Studies 53*, 251–267.
- Jackson, R., & Crawley, R. (2003). White student confessions about a black male professor: A cultural contracts theory approach to intimate conversations about race and worldview. *The Journal of Men’s Studies, 12*, 25– 41
- Johnson, W. L., Rickel, J. W., & Lester, J. C (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *Journal of Artificial Intelligence in Education 11*, 47-78.
- Jones, E., Farina, A., Hastorf, A., Markus, H., Miller, D., & Scott, R. (1984). *Social stigma: The psychology of marked relationships*. New York, NY: Freeman
- Kate (version 1) [Computer Software]. Santa Clara, CA: NeoSpeech
- Kierstead, D., D’Agostino, P., & Dill, H. (1988). Sex role stereotyping of college professors: Bias in students’ ratings of instructors. *Journal of Educational Psychology, 80*, 342–344.
- Lee, E.-J., & Nass, C. (1998). Does the ethnicity of a computer agent matter? An



- experimental comparison of human-computer interaction and computer-mediated communication. *Proceedings of the WECC Conference, Lake Tahoe, CA.*
- Lester, J., Conerse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY: AMC.
- Lester, J., Stone, B., & Stelling, G. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction, 9*(1-2), 1-44
- Mayer, R. E. (2001). *Multimedia learning.* Cambridge, England: Cambridge University Press.
- Microsoft Agent (Version 2) [Computer Software], Redmond WA: Microsoft
- Millis, K, Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. (2011) Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & J. Lakhmi (Eds.) *Serious games and edutainment applications* (pp.169-196). London, UK: Springer-Verlag
- Mimic Pro (Version 2) [Computer Software]. Draper, UT: DAZ
- Moreno, R., & Flowerday, T. (2005). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31*(2), 186-207.
- Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social

agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19(2), 177-213.

Nass, C., & Moon, Y., (2000). Machines and mindlessness: Social responses to computers. *Journal Social Issues*, 56, 81–103.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology*, 27, 10, 864-876.

Nass, C., Steuer, J., & Tauber, E.R. (1994). Computers are social actors. *Proceedings of Human Factors in Computing Systems*. Boston, Massachusetts.

Olney, A., D’Mello, S., Person, N., Cade, W., Hayes, P., Williams, C., Lehman, B., & Graesser, A.C. (2012). Guru: A computer tutor that models expert human tutors. In S. A. Cerri & B. Clancey (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems (ITS 2012)* (pp. 256-261). Berlin: Springer-Verlag.

Paivio, A (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press

Paul (Version 1) [Computer Software]. Santa Clara, CA: NeoSpeech

People Putty (Version 1) [Computer Software], Watsonville, CA: Haptik

Poser (Version 5) [Computer Software]. Aliso Viejo, CA: Smith Micro Software

Pratt, J., Hauser, K., Ugray, Z., Patterson, O. (2007). Looking at human-computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers*, 19(4), 512-523

- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media as real people and places*. Cambridge: Cambridge University Press/CSLI.
- Reid, L. D. (2010). The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com. *Journal of Diversity in Higher Education*, 3, 3.
- Ryu, J., & Baylor, A. L. (2005). The Psychometric Structure of Pedagogical Agent Persona. *Technology, Instruction, Cognition & Learning*, 2(4), 291-315
- Pratt, J., Hauser, K., Ugray, Z., & Patterson, O. (2007). Looking at human-computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers*, 19, (4), 512-523.
- U.S. Census Bureau. (2010). QuickFacts. Retrieved from <http://quickfacts.census.gov/qfd/>.
- Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2008). How to optimize learning from animated models. A review of guidelines based on cognitive load. *Review of Educational Research*, 78, 645-675.


## Appendix A

For \$7.25 an hour

Time (min)	Pay
1	\$0.12
5	\$0.60
10	\$1.21
15	\$1.81
20	\$2.42
25	\$3.02
30	\$3.63
35	\$4.23
40	\$4.83
45	\$5.44
50	\$6.04
55	\$6.65
60	\$7.25

## Appendix B

**The goal of this task is to assess the perceived ethnicity and gender of the character below.**



1. Please rate the agent's appearance on the level of Caucasian-ness.

- Very Caucasian
- Primarily Caucasian
- Undecided - Not more Caucasian than Non-Caucasian
- Undecided - Not more non-Caucasian than Caucasian
- Primarily Non-Caucasian
- Very Non-Caucasian

2. Please rate the agent on the level of Frisability?

## Appendix C

You are invited to participate in a research study being conducted at the University of Memphis by a graduate student (Kyle R. Cheney/901-406-5512/krcheney@gmail.com).

The purpose of this study is to examine users' preference for animated characters designed to facilitate learning in intelligent tutoring systems. It is estimate that it will take about one to three minutes of your time to complete an individual task. You are not permitted to repeat the tasks. You are free to contact the investigator at the above address and phone number to discuss the task.

Risks to participants are considered minimal. There will be no costs for participating. Any personally identifiable information will be stripped from the dataset prior to the researcher receiving it. Your responses will only be associated with the personal identification number assigned to you by the Mechanical Turk system and your information will remain completely anonymous to the researchers. Responses may be published, shared with other researchers, and/or presented in other Mechanical Turk questions to other participants.

Your participation in this survey is voluntary and you have the right to withdraw from participation at any time without penalty. If you have any questions, contact the investigator listed above.

This study has been reviewed and approved by The University of Memphis Institutional Review Board. If you have questions about your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact - anonymously, if you wish - the Institutional Review Board by phone at (901) 678-2533 or email at [irb@memphis.edu](mailto:irb@memphis.edu).

IRB Protocol Number: 2213

Please indicate if you agree to participate below.

Thank you.

I agree to participate in this study.  
 I do NOT agree to participate in this study.

## Appendix D

Hello, and 🙏 welcome!

My name is Professor Smith.

I'll be your professor today.

Together we will work through a short lesson.

## Appendix E

- **Please rate the agent's appearance on the level of Caucasian-ness.**

- 1 - Very Caucasian
- 2 - Primarily Caucasian
- 3 - Undecided - But more Caucasian than Non-Caucasian
- 4 - Undecided – But more non-Caucasian than Caucasian
- 5 - Primarily Non-Caucasian
- 6- Very Non-Caucasian

- **Please rate the agent on the level of femininity.**

- 1 - Very Feminine
- 2 - Primarily Feminine
- 3 - Undecided - But more Feminine than Non-Feminine
- 4 - Undecided – But more non-Feminine than Feminine
- 5 - Primarily Non-Feminine
- 6- Very Non-Feminine

- **Please rate the agent on the level of masculinity.**

- 1 - Very Masculine
- 2 - Primarily Masculine
- 3 - Undecided - But more Masculine than Non-Masculine
- 4 - Undecided – But more non-Masculine than Masculine
- 5 - Primarily Non-Masculine
- 6- Very Non-Masculine

- **Please rate the agent on the level of African American-ness.**

- 1 - Very African American
- 2 - Primarily African American
- 3 - Undecided - But more African American than Non- African American
- 4 - Undecided – But more non- African American than African American
- 5 - Primarily Non- African American
- 6- Very Non- African American



## Appendix F

1. Select your ethnicity from the list below:

- a. Caucasian
- b. African American
- c. Hispanic
- d. Asian
- e. Other

2. What is your age?

\_\_\_\_\_

3. What is your gender?

- a. Male
- b. Female

4. How many hours per week do you play video games?

\_\_\_\_\_

## Appendix G

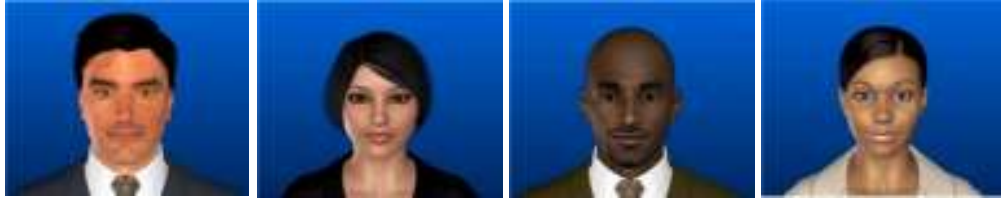
1. The agent led me to think more deeply about the training.
2. The agent made the training interesting.
3. The agent encouraged me to think about what I was learning.
4. The agent kept my attention.
5. The agent presented the material effectively.
6. The agent helped me to concentrate on the presentation.
7. The agent helped me focus on the relevant information.
8. The agent improved my knowledge of the content.
9. The agent was interesting.
10. The agent was enjoyable.
11. The agent seemed knowledgeable.
12. The agent seemed intelligent.
13. The agent was useful.
14. The agent was helpful.
15. The agent was instructor-like.
16. The agent had a personality
17. The agent's emotion was natural.
18. The agent was human-like.
19. The agent's movement was natural.
20. The agent showed emotion.
21. The agent was expressive.
22. The agent was enthusiastic.
23. The agent was entertaining.
24. The agent was motivating.
25. The agent was friendly.

## Appendix H

1. The agent was nice.
2. The agent was attractive.
3. The agent was warm.

## Appendix I

Which of the following Agents would you like to be your professor in a learning session?



1

2

3

4

Table 1

*The Average Score for all Four Agents on Each of the Four Validation Measures.*

Agent	African American-ness	Caucasian-ness	Femininity	Masculinity
CM	$M = 1.21$ $SD = .69$	$M = 5.35$ $SD = .70$	$M = 1.83$ $SD = 1.19$	$M = 5.05$ $SD = 1.28$
CF	$M = 1.83$ $SD = 1.41$	$M = 4.35$ $SD = 1.12$	$M = 5.4$ $SD = .98$	$M = 1.57$ $SD = .90$
AAM	$M = 5.62$ $SD = .77$	$M = 1.32$ $SD = 1.04$	$M = 1.47$ $SD = 1.17$	$M = 5.38$ $SD = .92$
AAF	$M = 5.25$ $SD = .70$	$M = 1.73$ $SD = 1.10$	$M = 5.75$ $SD = .53$	$M = 1.52$ $SD = 1.06$

*Note.* AAF: African American Female; AAM: African –American Male; CM: Caucasian Male; CF: Caucasian Female.

Table 2

*Pooled and Weighted Values and Significance Levels for the Four Similar and 12 Dissimilar Means.*

Variables	Similar	Dissimilar	<i>F</i> statistic
Human-Like	<i>M</i> = 3.72 <i>SD</i> = 1.20	<i>M</i> = 3.36 <i>SD</i> = .95	<i>F</i> (1, 109) = .50, <i>p</i> > .05
Credible	<i>M</i> = 3.01 <i>SD</i> = 1.01	<i>M</i> = 3.16 <i>SD</i> = .97	<i>F</i> (1, 109) = .01, <i>p</i> > .05
Engaging	<i>M</i> = 3.50 <i>SD</i> = 1.03	<i>M</i> = 3.55 <i>SD</i> = .85	<i>F</i> (1, 109) = 3.46, <i>p</i> > .05
Facilitates Learning	<i>M</i> = 3.50 <i>SD</i> = 1.10	<i>M</i> = 3.52 <i>SD</i> = .87	<i>F</i> (1, 109) = .01, <i>p</i> > .05
Affable	<i>M</i> = 3.03 <i>SD</i> = 1.00	<i>M</i> = 3.10 <i>SD</i> = .72	<i>F</i> (1, 109) = .02, <i>p</i> > .05

Table 3

*The Mean Score for the Measure of Human-Like.*

Participants		Agents			
		Caucasian		African American	
		M	F	M	F
Caucasian	M	$M = 3.66$ $SD = 1.11$	$M = 3.42$ $SD = .94$	$M = 3.54$ $SD = 1.03$	$M = 3.78$ $SD = 1.11$
	F	$M = 4.06$ $SD = 1.25$	$M = 4.02$ $SD = 1.12$	$M = 3.54$ $SD = 1.25$	$M = 3.89$ $SD = 1.30$
African American	M	$M = 3.77$ $SD = 1.21$	$M = 3.79$ $SD = 1.19$	$M = 3.40$ $SD = .1.21$	$M = 3.75$ $SD = 1.34$
	F	$M = 4.02$ $SD = 1.28$	$M = 4.22$ $SD = 1.18$	$M = 3.79$ $SD = .1.21$	$M = 3.80$ $SD = 1.23$

*Note.* M = Male; F = Female.

Table 4

*The Mean Score for the Measure of Engaging.*

Participants		Agents			
		Caucasian		African American	
		M	F	M	F
Caucasian	M	$M = 3.54$ $SD = .95$	$M = 3.28$ $SD = .87$	$M = 3.44$ $SD = .93$	$M = 3.55$ $SD = .87$
	F	$M = 3.71$ $SD = .97$	$M = 3.37$ $SD = 1.04$	$M = 3.16$ $SD = .99$	$M = 3.54$ $SD = 1.10$
African American	M	$M = 3.58$ $SD = 1.05$	$M = 3.61$ $SD = .95$	$M = 3.08$ $SD = .98$	$M = 3.21$ $SD = 1.29$
	F	$M = 3.94$ $SD = 1.04$	$M = 3.25$ $SD = 1.18$	$M = 3.25$ $SD = 1.01$	$M = 3.65$ $SD = 1.02$

*Note.* M = Male; F = Female.



Table 5

*The Mean Score for the Measure of Credibility.*

Participants		Agents			
		Caucasian		African American	
		M	F	M	F
Caucasian	M	$M = 3.04$ $SD = 1.16$	$M = 3.09$ $SD = 1.16$	$M = 3.16$ $SD = 1.08$	$M = 3.43$ $SD = 1.20$
	F	$M = 3.15$ $SD = 1.13$	$M = 3.16$ $SD = 1.09$	$M = 2.80$ $SD = .93$	$M = 3.10$ $SD = 1.19$
African American	M	$M = 3.14$ $SD = .92$	$M = 3.10$ $SD = .74$	$M = 2.75$ $SD = .69$	$M = 2.90$ $SD = .82$
	F	$M = 3.34$ $SD = .91$	$M = 3.47$ $SD = .84$	$M = 3.15$ $SD = 1.15$	$M = 3.08$ $SD = .89$

*Note.* M = Male; F = Female.

Table 6

*The Mean Score for the Measure of Affability.*

Participants		Agents			
		Caucasian		African American	
		M	F	M	F
Caucasian	M	$M = 3.18$ $SD = 1.08$	$M = 2.76$ $SD = .87$	$M = 3.38$ $SD = .89$	$M = 3.16$ $SD = 1.11$
	F	$M = 3.16$ $SD = .88$	$M = 3.03$ $SD = 1.00$	$M = 2.77$ $SD = 1.04$	$M = 3.02$ $SD = 1.01$
African American	M	$M = 3.45$ $SD = 1.11$	$M = 3.12$ $SD = .98$	$M = 2.94$ $SD = 1.04$	$M = 2.78$ $SD = 1.07$
	F	$M = 3.18$ $SD = .96$	$M = 3.20$ $SD = 1.06$	$M = 2.79$ $SD = 1.06$	$M = 2.96$ $SD = .86$

*Note.* M = Male; F = Female.

Table 7

*The Mean Score for the Measure of Facilitates Learning.*

Participants		Agents			
		Caucasian		African American	
		M	F	M	F
Caucasian	M	$M = 3.40$ $SD = 1.04$	$M = 3.24$ $SD = .99$	$M = 3.41$ $SD = .96$	$M = 3.64$ $SD = 1.00$
	F	$M = 3.66$ $SD = 1.03$	$M = 3.67$ $SD = 1.12$	$M = 3.22$ $SD = .85$	$M = 3.66$ $SD = 1.23$
African American	M	$M = 3.46$ $SD = 1.08$	$M = 3.48$ $SD = .91$	$M = 3.22$ $SD = 1.03$	$M = 3.09$ $SD = 1.07$
	F	$M = 3.81$ $SD = .95$	$M = 3.86$ $SD = .95$	$M = 3.60$ $SD = 1.07$	$M = 3.71$ $SD = 1.08$

*Note.* M = Male; F = Female.

Table 8

*The Results the Chi Square Test of Agent Choice by Participant Group.*

Participants	Observed	Expected	Residual
Caucasian Male Participant Ratings			
Caucasian Males	6	7.5	-1.5
Caucasian Females	16	7.5	8.5
African American Males	4	7.5	-3.5
African American Females	4	7.5	-3.5
Caucasian Female Participant Ratings			
Caucasian Males	8	6.5	1.5
Caucasian Females	6	6.5	-.5
African American Males	5	6.5	-1.5
African American Females	7	6.5	.5
African American male Participants Ratings			
Caucasian Males	2	6	-4
Caucasian Females	11	6	5
African American Males	8	6	2
African American Females	3	6	-3
African American female Participant Rating			
Caucasian Males	5	8.3	-3.3
Caucasian Females	9	8.3	.8
African American Males	9	8.3	.8
African American Females	10	8.3	1.8