The Priming Effects of Avatars in Virtual Settings

Jorge Peña,[1] Jeffrey T. Hancock,[2] and Nicholas A. Merola[1]

Abstract

The study extends research on the Proteus effect by demonstrating that avatars can prime negative attitudes and cognition in desktop virtual settings. Experiment 1 shows that, after virtual group discussions, participants using black-cloaked avatars developed more aggressive intentions and attitudes but less group cohesion than those using white-cloaked avatars. In Experiment 2, individual participants using a Ku Klux Klan (KKK)-associated avatar created more aggressive Thematic Apperception Test stories in comparison to a control group. Participants using the KKK avatar also wrote less affiliative stories in comparison to those employing avatars dressed as doctors. Overall, the resulting pattern of activation of negative thoughts (i.e., aggression) coupled with the inhibition of inconsistent thoughts (i.e., cohesion, affiliation) is consistent with principles of current priming models and provides initial evidence for automatic cognitive priming in virtual settings.

Keywords

Priming, automaticity, avatars, computer-mediated communication, Proteus effect, media effects, virtual environments

When people play video games and interact in virtual environments, they adopt social stereotypes and roles (e.g., soldier, doctor, mafioso, wizard) and interact in situations that go beyond real life (Shapiro, Peña, & Hancock, 2006). In doing so, people "become" someone else by employing digital bodies or *avatars* that serve as users' self-representation in a virtual setting (Eastin, 2006; Yee & Bailenson, 2007).

In most cases, the morals of video game characters are implied by the appearance and the stereotypical associations raised by their avatars (Isbister, 2006). For instance, one visual stereotype is that evil characters are portrayed with pale avatars dressed in dark clothes (e.g., "Arthas" in the lore of *Warcraft*), while good characters are usually depicted in light and vivid colors. But as we employ avatars to represent ourselves in computer-mediated

[1]The University of Texas at Austin
[2]Cornell University

Corresponding Author:
Jorge Peña, 1 University Station A 1105, Office 7.128, The University of Texas at Austin, Austin, TX 78712
Email: jorge.pena@mail.utexas.edu

settings, does our cognition change in relationship to our virtual persona? And if so, what mechanisms underlie the influence of avatars on users' cognition?

Initial studies have uncovered reliable evidence that avatars can affect users' behavior. Consistent with the prediction that attractiveness bestows more confidence in social interactions, Yee and Bailenson (2007) found that, in an immersive 3D environment, participants using avatars with more attractive faces walked closer and disclosed more information when compared to those using avatars with less attractive faces. In addition, consistent with the assumption that tallness confers higher status, participants using taller avatars tended to negotiate more forcefully in comparison to those using shorter avatars. Yee and Bailenson (2007) referred to the effects of virtual self-representations on users as the *Proteus effect*. Building on self-perception theory (Bem, 1972), Yee and Bailenson (2007) hypothesized that the Proteus effect is explained by people evaluating themselves as an imaginary third party would, observing their own behaviors to explain what attitudes may have caused them.

The present study extends research on the Proteus effect in several ways. First, we attempt to demonstrate that avatars can lead to effects on users' cognition in desktop in addition to virtual reality environments. Second, although the behavioral dimension of the Proteus effect has been examined from a self-perception standpoint, here we examine the cognitive dimension of the Proteus effect from a priming perspective. Based on this approach, situational cues should activate related thoughts (e.g., aggression) and inhibit inconsistent concepts (e.g., friendliness; Anderson & Spellman, 1995; Bargh, Chen, & Burrows, 1996; Dijksterhuis & van Knippenberg, 1996).

We also test whether more negative Proteus effects can be observed. Instead of manipulating avatar attractiveness to influence participants' friendliness or modify avatar height to instill more confidence as in Yee and Bailenson's study (2007), we investigate whether avatars in uniforms with aggressive cues instill more negative thoughts on users. This is accomplished by means of well-established appearance manipulations, including random assignment to groups in black uniforms (Frank & Gilovich, 1988) and asking individuals to wear Ku Klux Klan (KKK) outfits (Johnson & Downing, 1979). Congruent with the finding that black color is associated with death and evil (Adams & Osgood, 1973) and white color has more positive connotations including goodness and helpfulness (Meier, Robinson, & Clore, 2004), groups with participants wearing black uniforms chose more aggressive games to play (e.g., darts) than participants wearing white uniforms (e.g., block-building; Frank & Gilovich, 1988). Similarly, Johnson and Downing (1979) found that costumes and uniforms can operate as cues to aggression. In their study, participants wearing KKK costumes increased their electric shocks when in comparison to participants wearing nurse costumes.

Below we introduce the automaticity model, which is one current priming model explaining how situational cues, such as one's own avatar in a virtual setting, can affect cognition and behavior.

The Effects of Situational Cues on Cognition and Behavior

Priming occurs when people think, act, or feel in a manner consistent with situational cues without the intention to do so or the awareness of having done so (Bargh, 1994; Berkowitz,

1974, 1984). In this view, situational cues linked to memories and stereotypes increase the probability that a perceiver will act in accordance with those associations without conscious thought (Bargh et al., 1996).

The automaticity model is one recent theoretical framework that builds from and extends previous priming models. The automaticity model proposes that the perception of associations and stereotypes increases the readiness of cognitive and behavioral responses with varying degrees of awareness, intention, efficiency, and personal control (Bargh, 1994, 1996; Bargh & Chartrand, 1999). The model focuses on how such perception-reaction links become automatized, chronically accessible, and hard-wired with no necessary act of will (Bargh & Chartrand, 2000). The automatic effects of stereotypes and situational events are modeled on two elements: perception and behavior. In particular, the perception of situational and stereotypical features primes expectancies for what will happen, subjective evaluations of what is happening, emotional reactions one has had in related situations, and other cognitive, affective, and behavioral patterns (Bargh et al., 1996).

For example, priming the stereotype of professors among study participants increased their performance in a knowledge scale, but those primed with the stereotype of soccer hooligans performed worse (Dijksterhuis & van Knippenberg, 1998). Also, participants exposed to words implying rudeness later interrupted a conversation more often than participants who received the neutral prime, while those who received the politeness prime interrupted the least (Bargh et al., 1996, Experiment 1).

These predictions are built on the assumption that people's responses are mentally represented by knowledge structures. For example, Collins and Loftus' (1975) concept of *spreading activation* proposes that concepts (e.g., "KKK") are stored in memory as a node in a network of related information (e.g., "racism," "aggression," "burning crosses"). Therefore, priming one concept (e.g., "KKK") can also activate related concepts (e.g., "aggression") depending on the number of features held in common by the two concepts and how strong the relationship between concepts is. There is also evidence for the existence of *spreading inhibition* in relationship to knowledge constructs that are inconsistent with each other (Anderson & Spellman, 1995). For example, the activation of the stereotype of soccer hooligans also leads to the inhibition of friendliness (Dijksterhuis & van Knippenberg, 1996).

The automaticity model's implication for the Proteus effect is clear. Negative avatar stereotypes linked to aggression (e.g., avatars dressed in black uniforms or in KKK robes) will activate aggressive thoughts and, at the same time, will also inhibit more positive thoughts.

The Present Experiments

The automaticity model specifies conditions under which an avatar may affect users' cognition. These predictions are tested by means of adopting established appearance manipulations to a virtual setting. Rather than dressing participants in uniforms, they will employ avatars in black or white uniforms in Experiment 1 (Frank & Gilovich, 1988) and avatars with KKK and medical doctor associations in Experiment 2 (Johnson & Downing, 1979) using desktop computers. In this setup, the participants are fully in control of their avatar from a third-person visual perspective, and thus they simultaneously see and become the

character in the center of a computer screen (Figures 1 and 3). In these conditions, it is expected that employing aggressive avatar stereotypes should affect users' cognition by priming aggressive attitudes and thoughts while also inhibiting inconsistent thoughts (e.g., cohesion, affiliation). Thus, Hypotheses 1 and 2 predict the following:

Hypothesis 1 (H1): Participants employing avatars with negative associations will unintentionally display more negative thoughts.
Hypothesis 2 (H2): Participants employing avatars with aggressive associations will unintentionally display fewer positive thoughts.

Finally, it is important to note that the studies examining priming effects avoid making the link between the priming manipulation and the intended effect too obvious to participants (Bargh & Chartrand, 2000). According to Bargh (1994), awareness of the priming stimulus (e.g., "I used an avatar in a KKK robe") does not matter as long as the participants are not alerted of the intended influence of the manipulation (e.g., "Using an avatar in KKK robes was supposed to make me more aggressive"). However, in aggressive priming experiments, high levels of awareness about the aims of the study raise the possibility of mere role-playing and demand characteristic effects (see Page & Scheidt, 1971). More critically, such unwanted influences preclude clear demonstrations of priming effects (Bargh et al., 1996). According to Bargh and Chartrand (2000), if participants remain unaware of the intended effect of experimental manipulations, then one can be more confident in concluding that priming mechanisms are driving the observed empirical effect. Thus the present study uses these preestablished manipulation checks to elucidate participants' awareness levels.

Experiment 1: Priming Effects of Avatars in Black and White Uniforms

Method

Overview. Participants were randomly assigned to an avatar dressed in a black or white cloak and were asked to read a scenario. Participants then engaged in three-person group discussions of the scenario while employing black- or white-cloaked avatars. After group discussions, the participants completed attitude and group cohesion measures, and an awareness check.

Congruent with H1, participants using avatars in black uniforms will report significantly more aggressive intentions, attitudes, and subjective norms. As predicted by H2, participants using avatars in black uniforms will report less group cohesion in comparison to participants using avatars in white uniforms.

Participants

Fifty-one communication students from a large northeastern university participated in this experiment in exchange for course credit. Out of these, 34 were women and 17 were men,

Figure 1. Virtual group discussions with black- and white-cloaked avatars

and the majority of them were native English speakers. The mean age of the participants was 19.84 (SD = 2.61). Participants had spent ample time using personal computers (M = 10.16 years, SD = 2.85 years), but video game experience in the sample was lower on average (M = 4.68 years, SD = 5.74 years). All participants were assigned to three-person virtual groups (N = 17) and communicated exclusively via text messages sent within the virtual reality.

Materials

Avatars and setting. The avatars and the virtual setting were modeled using a video game titled "Jedi Knight II: Jedi Outcast." Participants employed either a black- or a white-cloaked avatar. Both avatars depicted the same older White man; the only variation was the character either donning a black or a white cloak. The black-cloaked avatar was intended to resemble a negative stereotype such as "executioner" or "evil Jedi," whereas the more positive white-cloaked avatar resembled a "monk" or "good Jedi" (see Figure 1). All weapons originally included in the game were disabled to avoid confrontation among participants.

Task. The task presented a scenario in which an online video game player repeatedly committed an offense (i.e., "attacking unarmed players"), and described five possible punishments for this behavior. Participants in virtual groups were asked to rank order the five punishments and come to group consensus on one punishment for the offender. The punishments varied in their intensity as they reflected mild (e.g., the offender is sentenced to apologize publicly) to strong punishments (e.g., the offending character is deleted/killed and can never be used again). These punishments were presented in a random order that was kept constant throughout the experiment. Participants were also assigned to a group or individual identity to examine how identity affected the decision-making process. These identity-related results are presented elsewhere (Merola, Peña, & Hancock, 2006). Note that this identity manipulation did not interact with any of the results presented here.

Design and Procedure

The experiment involved a 2 (Avatar: black vs. white uniform) × 2 (Gender: male vs. female) factorial design. Participants were recruited using the cover story of an "online interaction study" and were randomly assigned to the experimental conditions. Individual participants were led to testing rooms with a computer terminal and were never visually identifiable to each other. After granting informed consent, the participants filled in a demographic information survey measuring age, computer experience, and video game experience. After completing the demographic information survey, participants were shown mug-shot pictures of their randomly assigned black- or white-cloaked avatar before immersing in the online video game. The experimenter briefly mentioned to participants that this is who they would become in the virtual setting. Then participants were instructed on how to move the avatar and send text messages. Participants then proceeded to interact in three-person virtual groups in which all group members employed the same (black- or white-cloaked) avatar (see Figure 1). Recall that, in reality, individual participants were physically isolated in three networked rooms with desktop computers. Participants were given 50 minutes for discussion but group interactions usually lasted between 15 and 20 minutes. Discussions were stopped after consensus on a punishment was reached. Following this, participants individually completed attitude and group cohesion scales and an awareness check supplied by Page and Scheidt (1971).

Dependent Measures

Attitude and group cohesion measures were employed to assess avatar effects on participants' cognition.

Attitudes toward "attacking unarmed players." A questionnaire assessing attitudes toward breaking the norm depicted in the task (i.e., do not attack unarmed players or "laming") measured postdiscussion priming effects on a 7-point Likert-type scale ranging from 1 to 7. The attitude questions reflected principles stemming from the theory of planned behavior (Ajzen & Fishbein, 2000). In particular, the participants were asked to state if (a) I would "lame" someone if I ever played this game using this character (i.e., *behavioral intention*, extremely unlikely/extremely likely), (b) using this character, for me to commit "laming" on someone would be good (i.e., *attitude toward the behavior*, very bad/very good), (c) others looking like me would commit "laming" (i.e., *subjective norm*, extremely unlikely/extremely likely), and (d) if I was using this character, I could "lame" someone while playing the game (i.e., *perceived behavioral control*, definitely false/definitely true). The reliability of the 4-item attitude scale was acceptable ($\alpha = .70$). Reliability increased to ($\alpha = .78$) when the perceived behavioral control item was dropped. Therefore, the overall attitude model toward attacking an unarmed player only considered participants' intention, attitude toward the behavior, and subjective norm.

Group cohesion. Cohesion represents the shared beliefs and perceptions of individuals in a group and illustrates the strength of group identity (Seashore, 1954). The cohesion items were developed by Seashore (1954) and assessed dimensions of group cohesion such as

belongingness (i.e., To what degree did you feel that you are really a part of your group?), members' relationships (i.e., How does your group compare with other groups on the way people get along together?), and helping (i.e., How does your group compare with other groups on the way people help each other on the task?). The cohesion items were formulated in a 7-point Likert-type scale ranging from 1 to 7. Reliability for the three items in the cohesion scale was acceptable (α = .80).

Data preparation and statistical design. All dependent measures were examined using Tukey's 1.5 IQR criteria (Dollinger & DiLalla, 2006) to identify bivariate outliers (i.e., avatar by dependent variable). Each bivariate outlier was excluded from the pertinent data analysis and then returned to the total pool of 51 participants for subsequent tests because individual participants were not outliers on multiple dependent variables. The statistical analysis involved a linear mixed model approach since the data were obtained through nonindependent observations.

Results

Awareness Check. Based on procedures outlined by Page and Scheidt (1971), the participants were asked whether they were suspicious of the manipulations in the experiment, whether they could guess the research hypotheses, and if so to state what they thought the hypotheses were. Participants' responses showed that they were not suspicious about any experimental feature, including the role of the avatars. Most participants thought the experiment related to group decision making. Therefore, all of the participants were included in the analysis, as they did not appear to be aware about the intent of the experiment.

Intentions, Attitudes, and Subjective Norms for "Attacking Unarmed Players". Did participants' avatars affect their intentions, attitudes, and subjective norms related to upholding the standard depicted in the task (i.e., "do not attack unarmed players")? Consistent with H1, participants in the black-cloaked avatar condition stated a higher behavioral intention to attack unarmed players (M = 3.13, SE = 0.33) in comparison to participants in the white-cloaked avatar condition (M = 1.57, SE = 0.34), $F(1, 44)$ = 15.43, p < .001, η^2 = .44. Participants in the black-cloaked avatar condition also developed more aggressive attitudes toward the behavior (M = 3.04, SE = 0.34) in comparison to participants in the white-cloaked avatar condition (M = 1.83, SE = 0.35), $F(1, 44)$ = 8.17, p < .05, η^2 = .27. Finally, participants in the black-cloaked avatar condition reported a subjective norm more in support of attacking other players (M = 3.38, SE = 0.29) compared to those in the white-cloaked avatar condition (M = 3.04, SE = 0.29), F < 1, p = ns. No other main or interaction effects were observed.

Overall, the results supported H1. Participants employing black-cloaked avatars reported more aggressive intentions and attitudes in comparison to participants using white-cloaked avatars (see Figure 2). This implies that, in comparison to those using white-cloaked avatars, participants using a black-cloaked avatar developed a higher intention to attack unarmed players and viewed attacking unarmed players more positively.

Group Cohesion. The results indicate that the avatars affected participants' group cohesion perceptions, $F(1, 46)$ = 8.48, p < .05, η^2 = .26. Participants in the black-cloaked avatar
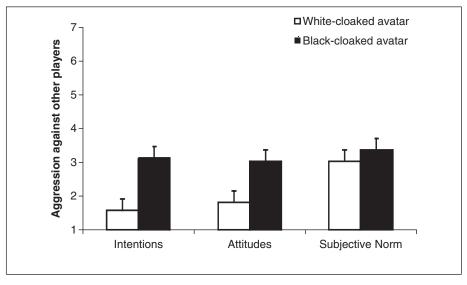
Figure 2. Means and standard error for aggressive intentions, attitudes, and subjective norms toward attacking other players by avatar condition

condition (M = 4.84, SE = 0.19) reported significantly lower group cohesion scores compared to participants in the white-cloaked avatar condition (M = 5.61, SE = 0.18). No other significant main effects were found on group cohesion scores. These results supported the inhibitory effect of aggressive-looking avatars anticipated in H2, as participants employing avatars with black uniforms displayed comparatively less cohesion.

The main effect was qualified by a significant interaction effect, $F(1, 14) = 17.67$, $p \leq$ .001, $\eta^2 = .78$. Decomposing the interaction effect revealed that male participants using black-cloaked avatars had significantly lower group cohesion scores (M = 4.67, SE = 0.27) than males using white-cloaked avatars (M = 6.25, SD = 0.27), $t(14) = 4.20$, $p <$ .001. The avatars did not significantly affect female participants' cohesion scores, $t(30) =$ 1.16, $p = ns$.

Discussion

The results of Experiment 1 lend support to the prediction that, in virtual settings, avatars with negative connotations affect users' cognition in line with the associations they raise. Participants employing black-cloaked avatars, of which "evil" and "aggression" are stereotypical attributes (Frank & Gilovich, 1988) developed significantly more aggressive intentions and attitudes than participants using white-cloaked avatars. Also, participants employing black-cloaked avatars reported significantly lower group cohesion scores in comparison to participants using white-cloaked avatars. The results are in line with H1 and H2, which predicted that avatars with aggressive associations should activate negative

cognitions and, at the same time, inhibit more positive cognitions. The manipulation check showed that participants remained unaware of the experiment's intent, suggesting that demand characteristics to respond pro or antisocially, or participants simply role-playing a part for the researchers, are unlikely to explain these effects.

How Do Avatars Affect Cognition? It is striking how simple avatar uniforms can affect participants' intentions, attitudes, subjective norms, and group cohesion levels. Since the participants remained unaware of influence, we feel more confident in concluding that priming mechanisms are related to the cognitive dimension of the Proteus effect (see Bargh & Chartrand, 2000).

However, Experiment 1 raises important questions. For example, although Frank and Gilovich (1988) specified that the effects of black uniforms is constrained to competing or deliberating groups, it is possible that having participants interacting in anonymous computer-mediated groups reinforced the norm implied by the avatars based on shared group identity. In support of this view, Lee (2004) found that participants attributed greater similarity to agents using same-looking avatars in virtual group discussions. Perhaps, having all virtual group members employ the same avatar enhanced the expected priming effect. This interpretation, however, cannot fully explain the pattern of results in Experiment 1. Although participants in both comparison groups used the same avatar, wearing either black or white-cloaks, participants using black-cloaked avatars reported lower cohesion (i.e., attraction to their group) than participants using white-cloaked avatars. This implies a differential effect based on the associations primed by uniform cues (see Johnson & Downing, 1979). Nonetheless, the second study addresses this potential shortcoming by assessing participants individually rather than in groups.

Also, participants' gender did not produce consistent results. Only men in black-cloaked avatars reported significantly lower group cohesion scores than men in white-cloaked avatars. The lack of gender effects might be linked to the fact that both male and female participants used the same male avatar. The second study overcomes this limitation by allowing participants to employ gender-matching avatars.

Experiment 2: Priming Effects of Avatars in KKK and Doctor Uniforms

Method

Overview. In Experiment 2, a KKK, a medical doctor, and a transparent avatar (a no-stereotype control condition) were used to prime the Proteus effect. What are the likely effects of priming KKK or doctor stereotypes by means of an avatar? Although personal encounters and media depictions of the KKK are less frequent nowadays in the United States, the KKK has strong negative connotations including aggression and intolerance (Johnson & Downing, 1979). In contrast, doctors are viewed far more positively in the United States. For example, TV series have linked the medical profession to attractiveness, achievement, power, knowledge, competence, and highly developed interpersonal skills and emotional stability (e.g., Pfau, Mullen, & Garrow, 1995). These studies suggest that the

stereotype of KKK members is consistently associated with aggression and other negative thoughts. Although medical professionals can sometimes raise negative connotations (e.g., Chory-Assad & Tamborini, 2003), the doctor stereotype is more positive than the KKK stereotype.

The cover story for participants was that it was a visual perception study in the form of a "virtual museum." Participants were told the experiment dealt with writing comments about pictures housed in the virtual museum. Individual participants wrote two stories based on Thematic Apperception Test (TAT) pictures while employing doctor, KKK, or transparent avatars. A postexperimental questionnaire included an awareness check and items representing alternative explanations.[1]

Consistent with H1's spreading activation assumptions, participants using KKK avatars, of which aggression, intolerance, and negativity are stereotypical attributes, will create the most aggressive TAT stories. Based on H2's spreading inhibition presuppositions, participants using the KKK avatar will compose stories with the least affiliation compared to participants using doctor and control avatars. Participants employing the KKK avatars will also compose TAT stories with the lowest use of achievement themes in comparison to the control and doctor avatars.

Participants

One hundred communication students from a large northeastern university were recruited in exchange for course credit. Six participants were disqualified from the experiment because they disobeyed or misunderstood the experimenter's instructions. Finally, 5 male and 1 female participant were excluded from the analyses because they showed high awareness about the purpose of the study (see manipulation checks). Therefore the final sample of Experiment 2 consisted of 38 men and 50 women ($N = 88$).

The average age of the participants was 20.32 years ($SD = 1.96$). In the final sample, 90.5% of the participants were native English speakers. Of these participants, 72.5% were Whites, 9.5% were Asian American or Pacific Islander, 9.5% were of African American descent, 6% were Latinos, 1.2% was American Indian or Alaskan, and 1.2% marked "Other" race. Similarly to Experiment 1, while the participants had ample years of experience with computers ($M = 11.59$, $SD = 2.49$), their video game experience was far lower. Only 10 participants devoted some weekly time to play video games, and many male participants expressed they stopped playing after getting into college.

Materials

Avatars and setting. Participants were randomly assigned to an avatar either resembling a doctor or a KKK member or to control avatars with a transparent body. The doctor and control avatars had male and female versions to match participants' gender. Notice that there was only one version of the KKK avatar because its gender was not evident under the hood and cloak. The avatars and the virtual setting were created using the video game "Jedi Knight II: Jedi Outcast." All weapons originally included in the game were removed.

The virtual museum has an entrance with a mirror where participants were introduced to their avatar. The corresponding avatar was always placed in front of the mirror by the entrance of the virtual museum. Next to the entrance, there was a large corridor and three more large mirrors at its end. The large mirrors at the right and left corners of the corridor were mirrored-doors that reveal the two rooms housing each TAT picture. Mirrors in virtual settings have been frequently used to reinforce avatar appearance to the participants (Eastin, 2006; Yee & Bailenson, 2007).

Task. The participants, while employing an avatar, were asked to create one story for each of the two TAT pictures housed in the virtual museum. Participants could only see one picture at a time since the images were in separate rooms within the virtual setting. In one picture (8BM of the TAT), "an adolescent boy looks out of the picture. The barrel of a rifle is visible at one side, and in the background is the dim scene of a surgical operation, like a reverie-image" (Murray, 1943, p. 19). Picture 8BM has been successfully employed to measure activation and inhibition of aggressive cognitions (Hokanson & Gordon, 1958).

The second picture in the virtual museum (17BM of the TAT) describes "a naked man clinging to a rope. He is in the act of climbing up or down" (Murray, 1943, p. 20). Overall, the emotional tone of stories elicited by this picture range from neutral to moderately happy and elicits story themes dealing with self-esteem, exhibition, competition with peers, escape from peril, curiosity, and aspirations (Eron, 1950).

Design and Procedure

Experiment 2 is a 3 (Avatar: doctor, control, KKK) × 2 (Gender: male vs. female) factorial design. After granting informed consent, the participants were asked to write one story for each of the two TAT pictures in the virtual museum. Participants were given standard TAT guidelines for creating their stories, which were placed in a sheet behind the participants' computer and were also read aloud by the experimenters. TAT researchers usually instruct their participants to write stories with a concrete plot (i.e., beginning, climax, conclusion) about what is happening in the picture, who these people are, what led to the situation they are in, what is being thought, what is wanted, by whom, what will happen, and what will be done (Atkinson, 1958; Murray, 1943). Participants were explicitly told to avoid coming up with several short stories from a single slide, composing stories with no outcomes, omitting characters and situations, rambling stories, and stories caught on detail with no clear plot.

The experimenter turned on the desktop computer's screen after explaining the guidelines to participants to reveal the randomly assigned avatar. The avatar was always presented in third person view and always in the same position by the mirror in the entrance of the virtual museum. As in Johnson and Downing's (1979) study, the appearance of the avatar was immediately presented as an accident of convenience by the experimenter (i.e., "I'm not much of a designer; this thing came out looking kind of _____ [like a Doctor/ Ku Klux Klannish/Transparent]"). The participants then received instructions on how to move the avatar and send text messages.

Figure 3. TAT picture 8BM (boy and surgery) from the perspective of male participants using avatars in doctor, control, and KKK uniforms

After the training session, the participants were asked whether they understood the guidelines of the writing task and whether they felt comfortable when operating their avatar. If that was the case, they were invited to stroll along the museum's corridor and enter one of the two rooms housing each TAT picture. The presentation order of the pictures was alternated by the experimenter, who told participants which of the two virtual rooms to enter first. The experimenter advised the participants to stay in a picture room until the story was completed to avoid stories merging the attributes of both pictures, comparisons between pictures, and the like.

The participants were given 15 minutes to compose each story after they entered a room with a TAT picture (see Figure 3). To avoid frustration with the experiment, the participants were also allowed to finish their stories voluntarily if they had been writing for 10 minutes and their story fulfilled the guidelines. Following the writing task, the experimenter told the participants the experiment was over. All TAT stories were electronically stored. Participants were then immediately invited to provide demographic information and respond to scales on alternative mechanisms, a questionnaire not reported here, and an awareness check supplied by Bargh and Chartrand (2000). Rotated sequences of the different scales were used to avoid order effects.

Dependent Measures

TAT stories. Avatar priming effect on cognition was operationalized as differences in the themes of the TAT stories created in each avatar condition. Two judges were trained to rate each story's theme using four unipolar scales ranging from 1 to 6. Each of the participants' stories were rated on how much its main theme contained *affiliation*, *aggression*, *achievement*, and *power* concerns, which are some of the most studied attributes in TAT research (Smith, 1992).

High TAT affiliation was reflected in story themes referring to establishing, maintaining, or restoring a positive affective relationship with another person or group (Smith, 1992). Stories that were high in TAT aggression had themes dealing with hate, anger, and feelings of ill will, where characters got involved in quarrels, cursing, reproving, blaming, ridiculing, or exciting aggression against another person (Murray, 1943). Stories that were high in TAT achievement had themes where the primary concern was competing with some

standard of excellence, winning or doing as well as (or better than) someone else, or meeting self-imposed standards of high performance (Smith, 1992). Finally, stories that were high in TAT power described plots about influencing people, asserting the self or overcoming someone's influence, where characters make explicit or implicit attempts to gain or maintain control, command, persuade, dominate, or convince someone (Smith, 1992).

The two judges were blind to the study's hypotheses and to participants' avatar and gender. The judges were trained to rate the themes using the discarded stories of participants with high awareness. Both judges rated all of the participants' stories. Judges' agreement was calculated using correlations since their ratings were a continuous rather than a categorical variable (Bargh et al., 1996; Frank & Gilovich, 1988). Judges' ratings for TAT affiliation ($r = .73$, $p < .001$), TAT aggression ($r = .82$, $p < .001$), and TAT achievement ($r = .81$, $p < .001$) were highly related and were included in the analysis. However, the TAT power ratings were substantially less related ($r = .51$, $p < .001$) and therefore were excluded from subsequent analyses.

Data preparation and statistical design. All dependent measures were scanned in search of bivariate outliers (i.e., avatar by dependent variable) using the 1.5 IQR criteria (Dollinger & DiLalla, 2006). Each bivariate outlier was excluded from the pertinent data analysis and then returned to the total pool of 88 participants for subsequent tests because individual participants were not outliers on multiple dependent variables. The statistical analysis involved a 3 (Avatar: doctor, control, KKK) × 2 (Gender: male, female) general linear model approach.

Results

Awareness check. Participants were asked about their awareness of the manipulations in the experiment, whether they could guess the research hypotheses, and if so to state the hypotheses (see Bargh & Chartrand, 2000). Six participants showed high awareness of the experiment's hypotheses and the priming function of the avatars and therefore were excluded from the analyses reported below. Most of these participants were men enrolled in video game design and information science courses. The remaining participants stated they could not guess what the experiment was about or believed the study dealt with one's personality and storytelling, how personal creativity influenced storytelling, or how one's liking of the virtual character affected stories or questionnaires, none of which are the main themes of this study.

Avatar effects on participants' TAT stories. Did doctor and KKK avatars affect participants' story themes? A general linear model analysis on the mean of the two judges' evaluations of affiliation in each story revealed a main effect of participants' avatar on the story ratings average, $F(2, 76) = 4.96$, $p < .05$, $\eta^2 = .12$. A post hoc Tukey test ($p < .001$) revealed that the story affiliation ratings of participants employing KKK avatars ($M = 1.93$, $SD = 0.79$) were significantly lower in affiliation than the stories from participants using doctor avatars ($M = 2.83$, $SD = 1.42$), and neither group differed from the control group ($M = 2.31$, $SD = 0.94$). The analysis also uncovered a main effect of participants' gender

on the thematic ratings, $F(1, 76) = 12.81$, $p < .001$, $\eta^2 = .14$. According to the raters, the stories of female participants included more affiliation themes ($M = 2.70$, $SD = 1.18$) in comparison to those of male participants ($M = 1.89$, $SD = 0.88$). No interaction effects between participants' avatar and gender were found on TAT affiliation ratings.

Themes of aggression in participants' TAT stories were significantly affected by the experimental avatars, $F(2, 77) = 5.05$, $p < .05$, $\eta^2 = .12$. Post hoc tests ($p \leq .05$) revealed a significant difference between TAT aggression themes in the KKK avatar condition ($M = 2.76$, $SD = 1.19$) in comparison to the control group ($M = 1.91$, $SD = 0.68$). The TAT aggression ratings in the doctor avatar condition ($M = 2.38$, $SD = 0.93$) did not differ from the KKK or control avatars, suggesting that only the KKK avatar primed aggression. No main effects of gender or interaction effects between avatar and gender were found on TAT aggression ratings.

Finally, although the doctor avatar condition had the highest achievement ratings ($M = 3.15$, $SD = 1.33$) relative to control ($M = 2.88$, $SD = 1.40$) and KKK conditions ($M = 2.74$, $SD = 1.44$), as predicted, these differences were not statistically significant $F(2, 76) < 1$. The stories of female participants were rated as higher in TAT achievement themes ($M = 3.26$, $SD = 1.30$) in comparison to male participants ($M = 2.46$, $SD = 1.38$), $F(1, 76) = 7.09$, $p < .05$, $\eta^2 = .09$.

Overall, the results indicate that the avatars primed differences in participants' TAT stories. In support of H1, participants employing the KKK avatar wrote stories with significantly higher levels of aggression in comparison to the control group (see Figure 4). H2 was partially supported: Participants using KKK avatars wrote stories with lower levels of affiliation in comparison to those using the doctor avatar, although there were no significant differences in story achievement ratings.

Discussion

Experiment 2's findings support the prediction that avatars with aggressive connotations can negatively affect users' cognition. In support of H1, participants that employed a KKK avatar composed TAT stories with more aggressive themes (i.e., murder, vengeance, crime, and scorn). Also, participants in KKK avatars created stories with lower levels of affiliation in comparison to those employing the doctor avatar, while the control group's affiliation level was between the doctor and the KKK avatar (see Figure 4). This lends some support to H2, which expected that using avatars with aggressive associations inhibit more positive thoughts. Overall, the results indicate that the negative activation and inhibition effect of the KKK avatar was more consistent than the effect of doctor avatars, echoing the finding that negative cues, situations, and events have stronger effects than more positive situations and outcomes (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001).

Experiment 2's manipulation checks confirmed that participants' remained unaware of how the avatars operated on the stories they wrote. This suggests that avatars had implicit, automatic effects on cognition as participants were not aware of the effect of their avatar. Perhaps surprisingly, only a few participants guessed the experimental hypothesis although the experiment presented a KKK avatar that could potentially raise suspicions or even
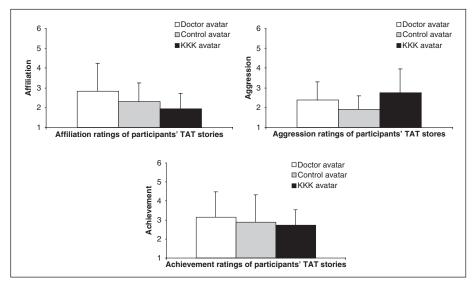
Figure 4. Means and standard deviations of judge's affiliation, aggression, and achievement ratings in TAT stories by avatar condition

offend the participants. The few participants who guessed the experimental hypotheses were enrolled in a class related to video games, giving some credence to Page and Scheidt's (1971) suggestion that experience with the subject is related to demand awareness. Finally, there were no reliable interaction effects between participants' gender and avatar type, suggesting that the avatars operated equally on men and women.

General Discussion

Across two experiments, avatars with negative associations elicited comparatively more negative thoughts on users and also inhibited positive concepts. Participants remained unaware of the influence of the avatars on their responses. The same priming techniques employed to assess how uniforms affect wearers' behavior (Frank & Gilovich, 1988; Johnson & Downing, 1979) produce similar effects when participants become such stereotypes in virtual settings by means of an avatar. The manipulation checks indicate that, in general, participants remained unaware of their avatar's influence on themselves and were not either role-playing or meeting experimenters' demands. Future studies should adopt similar control features when studying the effects of avatars on users' technologies to dispel doubts about awareness and role-playing effects over the findings.

Theoretical Implications

As noted earlier, previous research examining the behavioral aspects of the Proteus effect have relied on self-perception theory as an explanation for the effect (Yee & Bailenson,

2007). Although we have examined the cognitive aspects of the Proteus effect through the lens of priming and the automaticity model, can self-perception theory also account for the observed effects on cognition and attitudes? It seems clear that self-perception theory can potentially explain the aggressive attitudes and thoughts instilled by the negative appearance manipulations in the present study (e.g., Frank & Gilovich, 1988). Self-perception theory, however, seems less able to explain the inhibition of positive thoughts (e.g., cohesion, affiliation). Bem (1972) proposed self-perception theory as an explanation for how people come to know their attitudes based on past behavior, whereas the automaticity model's spreading activation and inhibition assumptions addresses not only how people come to know concepts but also how inconsistent thoughts are inhibited, deactivated, and even forgotten in the process (Anderson & Spellman, 1995; Dijksterhuis & van Knippenberg, 1996).

Although additional research is required, we argue that the cognitive effects observed in the present study are most parsimoniously explained by priming mechanisms. The present studies are inconclusive, however, in regard to the conditions under which self-perception and priming mechanisms make a relative contribution to the Proteus effect or when these mechanisms act independently. Future studies should specify under which conditions priming and self-perception mechanisms are independent or interdependent processes behind the Proteus effect.

Finally, future studies should disentangle the unique contribution of avatar appearance and role labels. One limitation of the present research is that, in emulating Johnson and Downing's (1979) study, Experiment 2 used both appearance and role manipulations. This procedure does not distinguish between the unique contribution of avatar appearance and role labels, and we are not investigating how avatar appearance and role labels contribute to the Proteus effect.

Social Implications

The associations raised by our online persona can affect users' cognition in virtual settings. Although avatars may elicit more positive outcomes such as increased friendliness and confidence (Yee & Bailenson, 2007) and augmented perceptions of similarity and identification (Lee, 2004), they can also activate and inhibit specific thoughts as shown in the present study. These outcomes depend on the attributes and roles highlighted by avatar design (Isbister, 2006). This effect should manifest itself in a variety of virtual settings in which people self-represent by means of avatars (e.g., "Second Life," "IMVU," "World of Warcraft").

One social implication of the study is that policy makers may hope to motivate designers to embed more positive associations, situations, and stereotypes in games. Alternatively, they can attempt to make users more aware about potential influences of the games and roles they play in the hope of activating controlled or intentional processes that regulate priming effects (see Bargh, 1994). These alternatives are worth exploring because, as the present study shows, virtual self-representation can surreptitiously affect our thoughts and attitudes.

Acknowledgments

Declaration of Conflicting Interests

Funding

Note

1. Alternative mechanisms included identification with one's avatar ("I identify with my virtual character," "The virtual character represents me"; $\alpha = .69$), arousal caused by the experiment ("This experiment was exciting," "Overall, I became bored with the experiment"; $\alpha = .76$; Anderson & Bushman, 2002), and frustration ("The virtual reality system was frustrating," "The virtual reality system was easy to employ"; $\alpha = .70$; Anderson & Bushman, 2002). The results, however, showed that the experimental avatars did not affect the identification and frustration scores, $F < 1$. However, the arousal self-report of women ($M = 4.28$, $SD = 1.01$) was significantly higher than the arousal self-report of men ($M = 3.64$, $SD = 1.00$), $F(1, 78) = 7.33$, $p < .05$, $\eta^2 = .09$. This effect might be connected to the fact that most women expressed having no experience with video games like the one employed in Experiment 2. Based on these results, alternative explanations for the findings in terms of mediating effects of participants' identification with the avatars, frustration levels, or arousal during the experiment appear untenable as there is no evidence for these mechanisms.

References

Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, *4*, 135-156.

Ajzen, I., & Fishbein, M. (2000). Attitudes and the attitude–behavior relation: Reasoned and automatic processes. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. *11*, pp. 1-33). New York: John Wiley.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*, 27-51.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *102*, 68-100.

Atkinson, J. W. (1958). *Motives in fantasy, action, and society*. Princeton, NJ: Van Nostrand.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum.

Bargh, J. A. (1996). Automaticity in social psychology. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169-183). New York: Guilford Press.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462-479.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253-285). New York: Cambridge University Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). The automaticity of social behavior: Direct effects of trait concept and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230-244.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323-370.

Bem, D. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. *6*, pp. 1-62). New York: Academic Press.

Berkowitz, L. (1974). Some determinants of impulsive aggression: Role of mediated associations with reinforcements for aggression. *Psychological Review*, *81*, 165-176.

Berkowitz, L. (1984). Some effects of thoughts on anti- and prosocial influences of media events: A cognitive-neoassociation analysis. *Psychological Bulletin*, *95*, 410-427.

Chory-Assad, R. M., & Tamborini, R. (2003). Television exposure and the public's perceptions of physicians. *Journal of Broadcasting & Electronic Media*, *47*, 197-215.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

Dijksterhuis, A., & van Knippenberg, A. (1996). The knife that cuts both ways: Facilitated and inhibited access to traits as a result of stereotype activation. *Journal of Experimental Social Psychology*, *32*, 271-288.

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology*, *74*, 865-877.

Dollinger, S. J., & DiLalla, D. L. (2006). Cleaning up data and running preliminary analyses. In F. T. L. Leong & J. T. Austin (Eds.), *The psychology research handbook: A guide for graduate students and research assistants* (pp. 241-253). Thousand Oaks, CA: Sage.

Eastin, M. S. (2006). Video game violence and the female game player: Self- and opponent gender effects on presence and aggressive thoughts. *Human Communication Research*, *32*, 351-372.

Eron, L. D. (1950). A normative study of the thematic apperception test. *Psychological Monographs: General and applied*, *64*(9), Whole No. 315.

Frank, M. G., & Gilovich, T. (1988). The dark side of self- and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, *54*, 74-85.

Hokanson, J. E., & Gordon, J. E. (1958). The expression and inhibition of hostility in imaginative and overt behavior. *Journal of Abnormal and Social Psychology*, *57*, 327-333.

Isbister, K. (2006). *Better game characters by design: A psychological approach*. San Francisco, CA: Morgan Kaufmann.

Johnson, R. D., & Downing, L. L. (1979). Deindividuation and valence of cues: Effects on pro-social and antisocial behavior. *Journal of Personality and Social Psychology*, *37*, 1532-1538.

Lee, E.-J. (2004). Effects of visual representation on social influence in computer-mediated communication: Experimental tests of the social identity model of deindividuation effects. *Human Communication Research*, *30*, 234-259.

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, *15*, 82-87.

Merola, N. A., Peña, J., & Hancock, J. T. (2006, June). *Avatar color and social identity effects on attitudes and group dynamics in online video games*. Paper presented at the 56th annual meeting of the International Communication Association, Dresden, Germany.

Murray, H. A. (1943). *Thematic apperception test manual*. Cambridge, MA: Harvard University Press.

Page, M. M., & Scheidt, R. J. (1971). The elusive weapons effect: Demand awareness, evaluation apprehension, and slightly sophisticated subjects. *Journal of Personality and Social Psychology*, *20*, 304-318.

Pfau, M., Mullen, L. J., & Garrow, K. (1995). The influence of television viewing on public perceptions of physicians. *Journal of Broadcasting & Electronic Media*, *39*, 441-458.

Seashore, S. (1954). *Group cohesiveness in the individual work group*. Ann Arbor, MI: Institute for Social Research.

Shapiro, M. A., Peña, J., & Hancock, J. T. (2006). Realism, imagination, and narrative video games. In P. Vorderer & J. Bryant (Eds.), *Playing computer games: Motives, responses, and consequences* (pp. 275-289). Mahwah, NJ: Lawrence Erlbaum.

Smith, C. P. (Ed.). (1992). *Motivation and personality: Handbook of thematic content analysis*. New York: Cambridge University Press.

Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33*, 271-290.

Bios

Jorge Peña, PhD, Cornell University, is an assistant professor in the Department of Communication Studies at The University of Texas at Austin. His research focuses on cognitive, affective, and behavioral processes involved in online collaboration and play.

Jeffrey T. Hancock, PhD (Dalhousie University), is an assistant professor in the Department of Communication and in the Faculty of Computing and Information Science at Cornell University. His research focuses on social interactions mediated by information and communication technology, with an emphasis on how people produce and understand language in this context.

Nicholas A. Merola is a doctoral student in the Department of Communication Studies at The University of Texas at Austin. His research explores anonymity, stereotyping, language use, and other elements in computer-mediated interaction.