Original Article

How Realistic Should Avatars Be?

An Initial fMRI Investigation of Activation of the Face Perception Network by Real and Animated Faces

Thomas W. James,[1] Robert F. Potter,[2] Sungkyoung Lee,[3] Sunah Kim,[4] Ryan A. Stevenson,[5] and Annie Lang[2]

[1]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA, [2]The Media School, Indiana University, Bloomington, IN, USA, [3]Strategic Communication, Missouri School of Journalism, University of Missouri, Columbia, MO, USA, [4]DMC R&D Center, Samsung Electronics, Co., Ltd., Suwon, South Korea, [5]Department of Psychology, University of Toronto, Toronto, ON, Canada

Abstract. Increased interaction with characters in games and online necessitates a better understanding of how different characteristics of these agents impact media users. This paper investigates a possible neurological underpinning for a common research finding – namely, that animated characters designed to be comparatively more human, more real, and more similar to the people they represent elicit more positive self-reported evaluations. The goal of this study was to examine the extent to which these results might be due to differential processing of character features in brain networks recruited for face recognition. There is some evidence that parts of the face network may be specifically tuned for real human faces. An experiment was conducted where participants viewed photographs of faces of actual agents (humans and animals) or colored drawings of matched agents (cartoon humans and animals). Using functional magnetic resonance imaging (fMRI) to measure blood oxygen-level dependent (BOLD) activation in the whole brain and specifically in the face network, we investigated the variation in patterns of activation with human and animal faces that were more or less real. The results were consistent with previous reports that the core regions of the face network are sensitive to the humanness of faces. However, our results extended previous work by showing that regions of the core *and* extended regions of the face network – and some regions outside the network – were sensitive to realism, but only realism of *human* faces.

Keywords: facial realism, fMRI, fusiform gyrus, amygdala

Animated Characters and Realism

A great deal of research demonstrates that people respond to media and media characters as social actors (Nass & Steuer, 1993; Reeves & Nass, 1996). With greater frequency, however, the media characters with which people interact are not exact images of human beings. Today a media interaction is as likely to be watching an animated avatar navigate within a video game as it is watching "real" people interact in a primetime drama. Responding to this phenomenon, much research has been published exploring how audience members interact with avatars, with results suggesting that people also respond to avatars as social actors (Bailenson et al., 2005; Nowak & Biocca, 2003; Nowak & Rauh, 2005). Many theoretical approaches argue that because social actors compel attention and trigger automatic emotional, social, and evaluative responses (Nowak, Hamilton, & Hammond, 2009), the mere presence of an avatar increases interaction in mediated realities (Taylor, 2011).

With the increased presence of avatars as mediated characters, and data indicating their presence has cognitive and emotional effects, it is important to try to better to understand how different characteristics of these agents impact media users. Several studies have demonstrated that changes in avatar features and characteristics influence the direction of the evaluative responses (Lewis, Weber, & Bowman, 2008; Weibel, Stricker, Wissmath, & Mast, 2010). The more an avatar is realistic and human, the more it is seen to be credible, attractive, and similar to the evaluator. In addition, these evaluations of animated characters can extend to influencing evaluations of the person being represented by them and of co-occurring content (Nowak et al., 2009). Human representations that are high in realism are perceived as persuasive and more similar than nonhuman and less-real characters (Guadagno, Blascovich, Bailenson, & McCall, 2007; Nowak et al., 2009; Park & Chung, 2011). This paper examines the possibility that responses to humanness and realness begin with differences in how the face perception network in the brain processes human and animal faces that appear more or less real. Because we are just beginning to investigate the influence of the face network, we conceptually define realism in a simple binary manner; things are more real, the more they

DOI: 10.1027/1864-1105/a000156

look exactly like a real object. Operationally then, photographs of humans and animals are defined as more real whereas cartoon drawings of humans and animals are less real.

This rudimentary definition of realism is quite different from work exploring the concept of the *uncanny valley* (MacDorman & Ishiguro, 2006), a term that originated with Mori (1970) who theorized that people would feel greater rapport with robots the more human-like they looked – but only up to a certain threshold of realism. Mori predicted that at some point the robot begins to look so much like a human that "we start to find its non-human imperfections unsettling" (MacDorman, Green, Ho, & Koch, 2009, p. 695). Much of the work in this area focusses on the impact of fine gradations in variables affecting perceived realism and self-reported feelings of unease or "eeriness" as the robot or avatar moves closer to the uncanny valley (MacDorman et al., 2009). Although some studies have explored the uncanny valley neurologically (Chaminade, Hodgins, & Kawato, 2007), they have not particularly focused on the face processing network as we do here.

Face Processing Networks

It has been well established that faces are processed in a broad network of brain regions – the *face network.* The *core* regions of the network include the fusiform face area (FFA; e.g., see Kanwisher, McDermott, & Chun, 1997; Kanwisher, Stanley, & Harris, 1999; Tong, Nakayama, Moskovitch, Weinrib, & Kanwisher, 2000), which responds to invariant aspects of faces, such as identity, and the superior temporal sulcus (STS) which responds to variable aspects of faces, such as expression. The *extended* regions of the network are so called because they are not involved in face perception/identification, but rather involved in extracting higher order meaning from faces (Haxby, Hoffman, & Gobbini, 2000; Rossion, 2008). Examples of these regions include the intraparital sulcus (spatial attention), auditory cortex (speech perception), limbic system (emotion), and anterior temporal cortex (biographical memory). The distributed face network – a core and extended conceptualization – is particularly interesting in terms of processing cartoon representations since separate brain regions within the network contribute to different cognitive, social, and affective operations specific to processing facial characteristics (Haxby et al., 2000; Rossion, 2008). Such operations may be differentially employed by virtual agents in different types of mediated interactions with users.

While it is clear that the humanness and realness of a face have a substantial influence on people's evaluative responses, the influences of humanness and realism on patterns of activation in the core and extended regions of the face network are equivocal. For example, some data suggest that the face network is *more* activated by real human faces than by either real animal faces or schematic human faces (Kanwisher et al., 1999; Tong et al., 2000). Other data, however, show that both cartoon faces and inverted cartoon faces activate the face network to the same extent as do photographs of real human faces (Tong et al., 2000). This could be because the face network is tuned to faces as a very broad class of objects.

The current study explored the extent to which the effects of humanness and realism on people's responses to neutral faces in mediated messages were determined by patterns of brain activation in the whole brain and specifically in regions of the face network. To that end, functional magnetic resonance imaging (fMRI) was used to measure blood oxygen-level dependent (BOLD) activation in the whole brain while participants viewed real human and real animal faces, and cartoon human and cartoon animal faces. Similar to that of Haxby et al. (2000), our study design also included a functional "localizer" task to identify the core and extended regions of the face network (for an accessible introduction to fMRI methodology, see Weber, Mangus, & Huskey, 2015). The key question was whether the human cognitive system would process the cartoon faces differently from how it does photographs of real human faces. In other words, is a photo of a face processed more like a face than a drawing of a face? And, secondarily, does it matter if that face is a human or an animal?

Method

Participants

Twenty-one healthy adults (11 women, ages 19–35) participated for payment. Two participants were excluded due to motion artifacts in the functional imaging data ($N = 19$). All research was performed in accordance with institutional review board approval, and participants signed informed consent forms prior to participation.

Stimuli

Two different sets of face stimuli were used, one set for the independent functional localizer run and a second set for the experimental runs. Face stimuli for the localizer were taken from the Max Planck Institute Face Database. These face stimuli are two-dimensional (2D) renderings of data derived from three-dimensional laser scans of human heads. To protect the confidentiality of the models, distortion (noise) was added to the data, such that the 2D rendered database face did not resemble the subject of the 3D scan. Thus, the face stimuli used in the localizer run were synthetic, but of such extremely high quality that they appeared to be photographs of real people's faces (Figure 1). All face stimuli used in the localizer were rendered at a 45″ view in grayscale. The comparison stimulus class for the localizer was grayscale photographs of houses. Employing this type of localizer procedure allowed for the precise functional specification of the various regions of the face network in this specific group of participants, rather than relying on published coordinates of regions developed from normed

Figure 1. Example of the face and house localizer stimuli.

data across many subjects. This paradigm has been used in many studies to localize the face network for subsequent region of interest (ROI) analyses (Poldrack, 2007; Saxe, Brett, & Kanwisher, 2006). The particular face and house stimuli used in this run were grayscale pictures of male and female faces in three-quarter view and grayscale pictures of buildings that have been used previously to identify the face network in individuals (James, Huh, & Kim, 2010).

Face stimuli for the experimental runs were selected from a larger collection of color face images that were scanned from books or taken from the Internet. Four types of stimuli were selected representing the poles zoomorph/anthromorph and natural/cartoon-like described by Strafling, Fleischer, Polzer, Leutner, and Kramer (2010). Operationally these were photographic images of real human and animal faces, and cartoon depictions of human and animal faces. The images contained more elements than just the face. Most images included the neck and shoulders and some background elements. The collection contained faces with a range of facial expressions. All face stimuli were initially rated on emotional arousal, emotional valence, and familiarity by 30 raters who were not participants in the experiment. Face stimuli were divided into neutral, positive, and negative valence (Bradley, Cuthbert, & Lang, 1996) based on the rating values. For each of the 12 stimulus conditions (four types by three valences), eight examples were selected such that they met two criteria: (1) that the mean familiarity rating across the eight stimuli was matched across the 12 conditions and (2) that the mean arousal rating across the eight stimuli was matched across the four types, separately for the neutral, negative, and positive valences. The entire experimental protocol included presentations of positive, neutral, and negative stimuli. Here only the neutral valence conditions were analyzed to provide the most controlled test of the impact of animation and human likeness on the neural substrates involved the facial network. These 32 images are provided in Figure 2.

Scanning Session Procedures

Participants lay supine in the scanner bore with their head secured in the head coil by foam padding. Participants viewed stimuli through a mirror that was mounted above the head coil.[1] This allowed the participants to see the stimuli on a rear-projection screen (40.2 × 30.3 cm) placed behind them in the bore. Stimuli were projected onto the screen with a Mitsubishi LCD projector (model XL30U). The viewing distance from the mirror to the eyelid was 11.5 cm, and the distance from the screen to the mirror was 79 cm, giving a total viewing distance of 90.5 cm. When projected in this manner, the size of the entire stimulus image subtended approximately 9″ of visual angle.

The fMRI data collection protocol included two types of runs. The first was the functional localizer run, in which the localizer face and house stimuli were presented with the goal of identifying functionally specific brain activity locations of the face network for this particular groups of participants. For the localizer run, we used a standard paradigm in which participants fixated on the center of the screen while blocks (multiple trials) of face stimuli were presented interleaved by blocks of house stimuli. Six stimuli per block were presented for 1,500 ms each with an interstimulus interval (ISI) of 500 ms, producing a block time of 12 s. Stimulus blocks were separated by 12-s fixation blocks. There were 12 stimulus blocks and 13 fixation blocks in the single run, making the total run length 5 min.

The second type of run was designed to present the experimental stimuli, displaying six images of a single level randomly selected (with replacement) within each factor repetitively in a block. Each block was followed by a rest period and then followed by a block of a different stimulus type, with each image again randomly selected (with replacement). Each experimental image was presented for 2,500 ms with an ISI of 500 ms. Therefore, stimulus blocks were 18 s long and were separated by 12-s fixation blocks during which participants were instructed to look at a small "+" mark in the center of a gray background. Each experimental run contained 12 stimulus blocks and 13 fixation blocks, for a total run length of 6 min and 12 s. Across the four runs, there were a total of four stimulus blocks per stimulus type. Matlab 5.2 (http://www.mathworks.com) combined with the Psychophysics Toolbox (http://www.psychtoolbox.org; Brainard, 1997; Pelli, 1997) was used to present stimuli and record participant responses during the scanning session.

We were then able to contrast patterns of neural activation in response to each stimulus type observed in the experimental runs through whole-brain analyses and also more specifically within the face network identified in the localizer run. The extent and selectivity of some regions of the face network is controversial (Ishai, 2008; Wiggett & Downing, 2008). Here, we analyzed five brain regions from the face network (Fox, Iaria, & Barton, 2009) – namely,

---

[1] A complete discussion of the intricacies of experimental design for fMRI studies are beyond the scope of this paper. Readers are encouraged to see Weber et al. (2015).

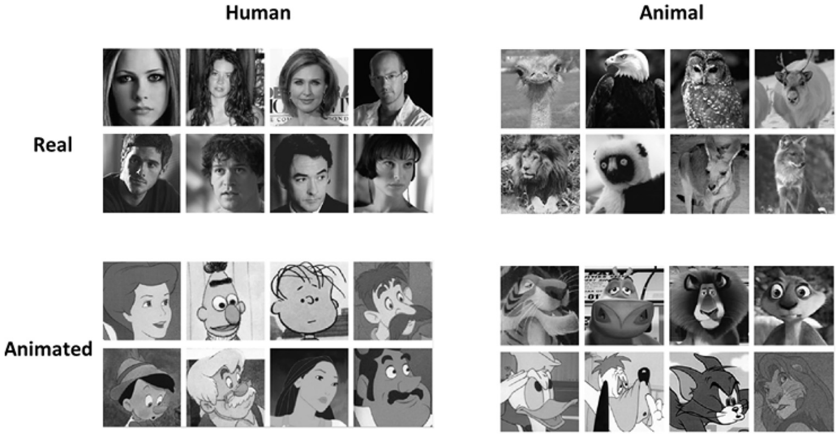


Figure 2. Stimulus images.

the FFA, the occipital face area (OFA), the STS, the amygdala (AMG), and the inferior frontal gyrus (IFG).

Imaging Parameters and Analysis

Imaging data were acquired with a Siemens Magnetom TRIO 3-T whole-body MRI scanner. Images were collected using an eight-channel phased-array head coil. The field of view was 220 × 220 mm, with an in-plane resolution of 64 × 64 pixels and 33 axial slices of 3.4 mm thickness per volume. These parameters produced voxels that were 3.4 × 3.4 × 3.4 mm. Functional BOLD images were collected using a gradient echo EPI sequence: TE = 25 ms, TR = 2,000 ms, flip angle = 70°. High-resolution T1-weighted anatomical volumes were acquired using a Turbo-flash 3-D sequence: TI = 1,100 ms, TE = 2.92 ms, TR = 2300 ms, flip angle = 12°, with 160 sagittal slices of 1-mm thickness, a field of view of 224 × 256 mm, and an isometric voxel size of 1 mm^3.

Imaging data were analyzed using BrainVoyager QX 2.2 (http://www.braininnovation.com). Individual anatomical volumes were transformed into a common stereotactic space based on the reference of the Talairach atlas using an eight-parameter affine transformation. This transformation puts each individual's neural activations into a common anatomical space, accounting for difference in brain size and shape. To account for motion during scans, all functional volumes were realigned to the functional volume collected closest in time to the anatomical volume using an intensity-based motion-correction algorithm. Functional volumes also underwent slice scan-time correction, 3D spatial Gaussian filtering (FWHM 6 mm), and linear trend removal. Functional volumes were coregistered to the anatomical volume using an intensity-based matching algorithm and normalized to the common stereotactic space using an eight-parameter affine transformation. During normalization, functional data were resampled to 3-mm^3 isometric voxels. Whole-brain statistical parametric maps were calculated using a general linear model (GLM) with predictors based on the timing protocol of the blocked stimulus presentation, convolved with a two-gamma hemodynamic response function. Beta weights were extracted from group ROIs using the ANCOVA Data Table tool in BrainVoyager's volume of interest module. Beta weights here represent the goodness of fit of activity within a particular region to a canonical hemodynamic response; thus higher beta weights indicate a greater level of neuronal activity in response to the presented stimulus condition. Statistical hypothesis testing was performed on the extracted beta weights using repeated measures ANOVAs in SPSS 19 (http://www.ibm.com/SPSS_Statistics). In the figures where graphs show error bars, those error bars represent 95% confidence intervals, calculated using the within-subjects mean squared error from the highest order interaction term.

Results

Data from two participants were removed from the analysis due to excessive motion during scans, therefore $N = 19$. The ROIs were determined from a group-average whole-brain fixed-effects GLM performed on the data from the independent functional localizer run. Group-average ROIs were localized by contrasting the face and house conditions with a corrected alpha of 0.05 (false discovery rate correction for multiple comparisons with $q = .05$) revealed significant clusters in regions of the extended face network, including the bilateral FFA, OFA, and IFG, the right STS and the right AMG (Figure 3; Table 1). Beta weights representing BOLD signal change were extracted from the ROIs for each participant. A three-way region by Humanness (human, animal) × Realism (photo, cartoon) (8 × 2 × 2) ANOVA with BOLD percentage signal change as the dependent variable revealed a significant three-way interaction, $F(7, 126) = 3.74$, $p = 0.001$.

As can be seen from Figure 4, the three-way interaction seemed largely due to the fact that four of the eight ROIs (Figure 4A; the right FFA, STS, AMG, and IFG) showed a relatively consistent pattern of differences across the four stimulus types, whereas the other four ROIs (Figure 4B; the bilateral OFA, and left FFA and IFG) showed few or no significant differences across the stimulus types. Separate three-way ANOVAs on these two groups of ROIs

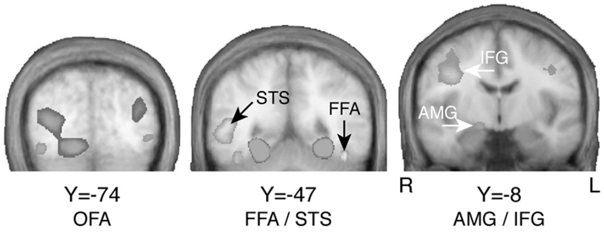


Figure 3. Locations of face-selective regions of interest. The face–object contrast is shown, thresholded with a *t* value of 6.0. Voxel regions identified by arrows indicate greater activation with faces. Y values are from the Talairach reference. AMG = amygdala; FFA = fusiform face area; IFG = inferior frontal gyrus; OFA = occipital face area; STS = superior temporal sulcus. Color figures available from authors.

Table 1. Talairach coordinates for regions of interest from functional localizer

Region	X	Y	Z
l-OFA	−43	−74	−1
r-OFA	+40	−74	−10
l-FFA	−42	−47	−14
r-FFA	+41	−45	−15
r-STS	+51	−52	+3
r-AMG	+15	−8	−7
l-IFG	−40	−5	+35
r-IFG	+42	+1	+29

Note. AMG = amygdala; FFA = fusiform face area; IFG = inferior frontal gyrus; l = left; OFA = occipital face area; r = right; STS = superior temporal sulcus.

supported the division. Neither ANOVA produced a significant three-way interaction, *F*(3, 54) = 1.64, *ns*; *F*(3, 54) = 1.08, *ns*, suggesting that the one-way and/or two-way patterns of activation across the four stimulus types were essentially the same within each group of four ROIs.

In the more "sensitive" group of four ROIs (the right FFA, STS, AMG, and IFG), the three-way ANOVA also showed a significant two-way interaction between Humanness and Realism, $F(1, 18) = 14.06$, $p < 0.001$, and a significant main effect of Realism, $F(1, 18) = 4.58$, $p = 0.03$. The two-way interaction effect, collapsed across the four regions, is shown in Figure 4C. The two-way interaction was due to a significant effect of Realism with human faces, but not with animal faces. Pairwise post hoc *t* tests confirmed that activation was higher with real human faces than cartoon human faces for all four ROIs in the "sensitive" group, all $t(18) > 2.66$, $p < 0.02$. In addition, a significant difference between real human and real animal face types was found for the FFA and AMG, both $t(18) > 4.25$, $p < 0.001$), but not for the STS or IFG.

In the "nonsensitive" group of four ROIs (the bilateral OFA, and left FFA and IFG), the three-way ANOVA showed only a significant two-way interaction between Region and Humanness, $F(1,18) = 3.48$, $p = 0.04$. This effect seemed to be driven by a strong preference of the left OFA for cartoon animal faces over the three other stimulus types, all $t(18) > 2.24$, $p < 0.04$. It is important to point out that overall signal change in these "nonsensitive" ROIs was at least as high as in the "sensitive" ROIs, suggesting that the lack of effects across stimulus types was not due to a general lack of activation to the stimuli in these regions.

To establish whether or not brain regions outside the face network were sensitive to realism, a whole-brain analysis was performed on the experimental runs using a GLM. The four stimulus conditions were modeled through time with four separate predictors created with delta functions based on the timing and order of the blocks in the experimental protocol. The time-based predictors were each convolved with a two-gamma hemodynamic response function. This is a standard step in a majority of fMRI analyses that provides a better fit to the data by mimicking the shape of the BOLD response. Estimates of the head motion produced by individual participants are sometimes added to the model as covariates (predictors of no interest), but they were not included in our model because motion was extremely small in all participants. Correction for the multiple testing problem was done using the Brainvoyager cluster-size threshold estimator plug-in. With a voxel-wise *p* value of .005 and a family-wise *p* value of .05, the required cluster threshold was five 3 × 3 × 3 mm voxels. A contrast of real and cartoon human faces is shown in Figure 5 using this combination of voxel-wise and cluster-size thresholds. Table 2 contains the coordinates of all of the significant clusters. As expected based on the results of the ROI analysis on the localized regions of the face network above, significant clusters were found in the fusiform gyrus (FG), STS, AMG, and IFG, although the posterior STS activation was in the left hemisphere instead of the right. Additional clusters were also seen in the anterior STS, the posterior occipital cortex, and the anterior insular cortex. A second contrast assessing the main effect of Realism across both human and animal faces did not show any significant clusters.

Discussion

As people spend more time in mediated interactions with avatars, other synthetic representations of people, and other synthetic anthropomorphized animals, it seems important for us to understand how the brain is responding differently to these more or less human and more or less realistic communication partners. This study sought to increase our understanding of how the brain and, in particular, brain systems for face perception, respond to these types of media stimuli. Here, BOLD fMRI was used to assess the sensitivity of the face network to the humanness (human vs. animal) and realism (real vs. cartoon) of neutral face stimuli. The results showed significant differences across the stimulus types. There was a strong preference for real over cartoon human faces in all regions of the

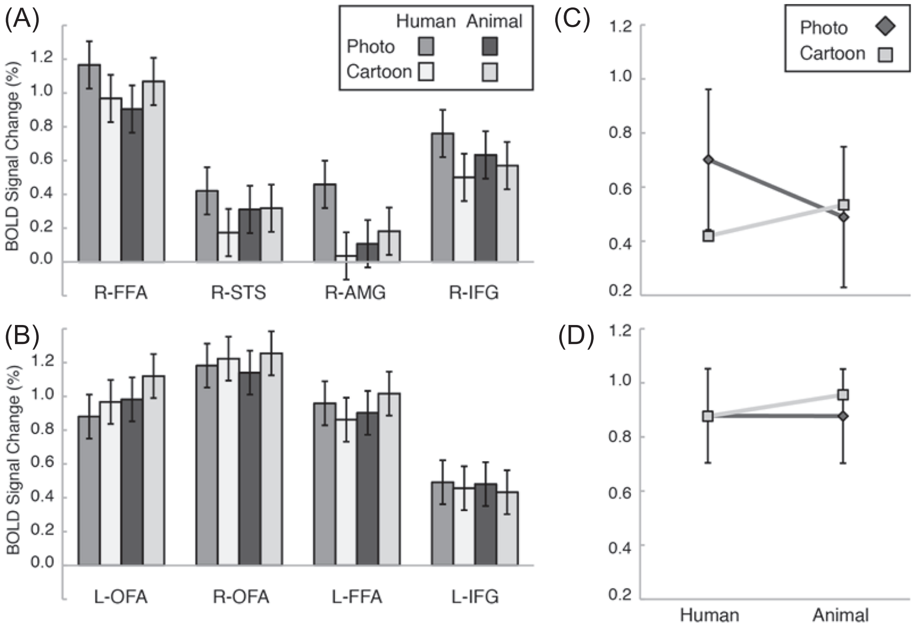


Figure 4. Blood oxygen-level dependent (BOLD) signal change as a function of brain region, humanness, and realism. Data from the four "sensitive" regions are shown in (A). Data from the other four regions are shown in (B). Two-way interactions collapsed across region are shown in the graphs on the right (C, D). Error bars are 95% confidence intervals.

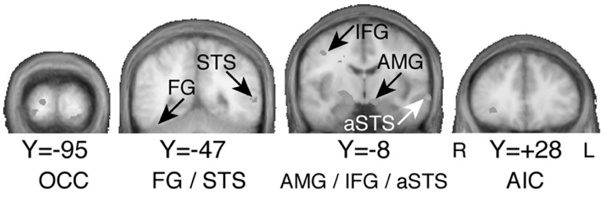


Figure 5. Statistical parametric map contrasting real and cartoon human faces. Voxel regions identified by arrows indicate greater activation with real faces. Y values are from the Talairach reference. OCC = posterior occipital cortex; FG = fusiform gyrus; STS = posterior superior temporal sulcus; AMG = amygdala; IFG = inferior frontal gyrus; aSTS = anterior superior temporal sulcus; AIC = anterior insular cortex. Color figures available from authors.

Table 2. Talairach coordinates for clusters from whole-brain analysis

Cluster	X	Y	Z
l-OCC	−18	−90	−11
r-OCC	+21	−97	−1
r-FG	+40	−41	−20
l-STS	+53	−44	+9
l-AMG	−10	−8	−9
r-AMG	+17	−8	−7
r-IFG	+40	−3	+34
l-aSTS	−61	−6	−6
r-AIC	+33	+28	+2

Note. AIC = anterior insular cortex; AMG = amygdala; aSTS = anterior superior temporal sulcus; FG = fusiform gyrus; IFG = inferior frontal gyrus; l = left; OCC = posterior occipital/peristriate; r = right.

right-hemisphere network except for the OFA. The previously documented preference in the right FFA, a core network region, for human over animal faces (Kanwisher et al., 1999) was replicated and extended to the AMG, but was not found in other areas. The results suggest that realism and humanness are influential factors when assessing brain activation in some – but not all – regions of the face network, as well as some regions outside of the face network.

Unlike the right FFA, the other core network region, the STS and OFA, did not show a preference for human over animal faces. However, both the posterior and anterior aspects of the STS did show a strong preference for real faces over cartoon faces, which is consistent with previous work (Mar, Kelley, Heatherton, & Macrae, 2007; Puce et al., 2003). Overall activation in the STS was quite low compared with the FFA and OFA, which is likely because the STS, as a region that is responsive to mutable aspects of facial processing such as expression, prefers stimuli in motion as opposed to the static stimuli used in the current study (Puce et al., 2003). Thus, it is possible that the lack of an effect of humanness was simply a floor effect caused by the lack of robust activation to the static stimuli. It is also possible that the STS would show tuning for humanness if human *movements* were compared with animal movements (Buccino et al., 2004; Mar et al., 2007). This might relate to research that suggests that embodied synthetic agents are evaluated most positively and have the most influence when they exhibit more realistic movements (Guadagno et al., 2007)

While the majority of face network regions in the right hemisphere showed a preference for real human faces over the other stimulus types, the right OFA did not. It has been suggested that the OFA contributes more to the early processing of face components than other regions of the face network (Betts & Wilson, 2010; Nichols, Betts, & Wilson, 2010; Pitcher et al., 2011). It may be that human and animal and real and cartoon stimuli do not differ as greatly at that level of processing. Higher levels of processing – which may involve analysis of features combined with analysis of configurations of those features and also analysis of the spatial relations among those features – may be more sensitive to differences between human and animal and real and cartoon face images.

Although a preference for real faces was found throughout the face network and beyond for human stimuli, no such effects were found for animal stimuli. This is in line with research that suggests that people are more likely to choose human over animal avatars and that they evaluate them more positively. Although realism is a salient dimension along which environmental stimuli can be easily categorized, and despite the fact that brain regions inside and outside the face network were highly sensitive to the realism of *human* faces, realism does not seem to be a general dimension along which the brain is sensitive for all classes of objects.

As virtual environments become more popular in our society (e.g., Second Life, Virtual World of Kaneva, Twinity), people are spending more and more time in them interacting with avatars. Although avatars were not specifically used in the current experiment, the results showed that the brain responds differently to cartoon human faces than to real human faces, suggesting that these differences may generalize to avatars as well. Given the role of the face network in a variety of face-related behaviors – recognizing identity, emotion, speech, intent – if such differences lead to poorer recognition, then virtual environments may turn out to be poorer environments for social exchange than anticipated. As our society trends toward the use of less real representations of humans in the media, it will be important to study the influence of these effects more thoroughly. We acknowledge that this is a single study and encourage replication. We also acknowledge limitations with this study. An interest in focusing on the questions of *whether* the concepts of realism and humanness influence the face network led us to be rather abrupt in our conceptual and operational definitions of both. It also led us to purposely not consider both motion and emotion in our experimental design and analytical approach. Avatars are rarely emotionally stoic or still. There is indeed substantial future work to be done to explore the fine details of the concepts of realism and humanness and their effects on activation of face networks. Certainly there are self-report results that suggest that synthetic/drawn human faces can reach a point of realism where the emotions presented are accurately identified (e.g. Dyck et al., 2008; Tinwell, Grimshaw, Nabi, & Williams, 2011). fMRI data suggests that this is even neurologically the case, with AMG activation in response to emotional avatar faces (Moser et al., 2007). The extent to which this begins with more robust activation of the face network in the brain remains to be investigated.

Those caveats aside, this study provides a baseline exploration into how mediated representations of characters are processed by the core and extended face network. Consistent with previous results, components of the core face network were sensitive to the humanness of faces in our data. Extending these results, many core *and* extended regions of the face network – and a few regions outside the face network – were sensitive to the realism of human faces. There are practical implications for these findings. Consider, for example, a game designer trying to decide whether to create a character that is realistically human or a cartoon. Along with genre and market demand considerations, the designer should consider the cognitive tasks the game player will employ when interacting with this character, and how easily they should be to execute to progress through the game at the desired pace. If allowing a game player to easily recognize the character within many visually busy scenes is an important goal, then our data suggest – due to greater activation in the core face networks – that the best way to accomplish this is with a comparatively more realistic, human character. Furthermore, if mere recognition is not all that is necessary, but a less taxing ability to recognize the expressions on the character's face is also required, then a picture-based human character once again would be suggested by our data which show it would lead to greater activation in extended brain regions associated with processing of facial expressions.

Acknowledgments

This research was supported in part by the Faculty Research Support Program, administered through the Indiana University Office of the Vice President of Research; and in part by the Indiana METACyt Initiative of Indiana University, in turn funded in part through a major grant from the Lilly Endowment Inc. We thank Karin Harman James for insights given into every aspect of the project and Samantha Brandfon for assistance with data collection.

References

Bailenson, J. N., Swinth, K. R., Hoyt, C. L., Persky, S., Dimov, A., & Blascovich, J. (2005). The independent and interactive effects of embodied agent appearance and behavior on self-report, cognitive, and behavioral markers of copresense in Immersive Virtual Environments. *PRESENCE: Teleoperators and Virtual Environments, 14*, 379–393.

Betts, L. R., & Wilson, H. R. (2010). Heterogeneous structure in face-selective human occipito-temporal cortex. *Journal of Cognitive Neuroscience, 22*, 2276–2288. doi: 10.1162/jocn.2009.21346

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1996). Picture media and emotion: Effects of a sustained affective context. *Psychophysiology, 33*, 662–670.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., … Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An FMRI study. *Journal of Cognitive Neuroscience, 16*(1), 114–126. doi: 10.1162/089892904322755601

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-actions. *Social Cognitive and Affective Neuroscience, 2*(3), 206–216.

Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., … Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PloS ONE, 5*(7), e11577.

Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, RC., et al. (2008). Recognition profile of emotions in natural and virtual faces. *PLoS ONE, 3*(11), e3628. doi: 10.1371/journal.pone.0003628

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping, 30*, 1637–1651.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & McCall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology, 10*(1), 1–22.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.

Ishai, A. (2008). Let's face it: It's a cortical network. *Neuroimage, 40*(2), 415–419. doi: S1053-8119(07)00988-3 [pii] 10.1016/j.neuroimage.2007.10.040

James, T. W., Huh, E., & Kim, S. (2010). Temporal and spatial integration of face, object, and scene features in occipito-temporal cortex. *Brain and Cognition, 74*(2), 112–122.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport, 10*(1), 183–187.

Lewis, M. L., Weber, R., & Bowman, N. D. (2008). "They may be pixels, buy they're MY pixels": Developing a metric of character attachment in role-playing video games. *Cyberpsychology and Behavior, 11*, 515–518. doi: 10.1089/cpb.2007.0137

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3), 695–710. doi: http://dx.doi.org/10.1016/j.chb.2008.12.026

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*, 297–337.

Mar, R. A., Kelley, W. M., Heatherton, T. F., & Macrae, C. N. (2007). Detecting agency from the biological motion of veridical vs animated agents. *Social Cognitive and Affective Neuroscience, 2*, 199–205. doi: 10.1093/scan/nsm011

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy, 7*(4), 33–35.

Moser, E., Derntl, B., Robinson, S., Fink, B., Gur, R. C., & Grammer, K. (2007). Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *Journal of Neuroscience Methods, 161*(1), 126–133. doi: http://dx.doi.org/10.1016/j.jneumeth.2006.10.016

Nass, C., & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research, 19*, 504–527.

Nichols, D. F., Betts, L. R., & Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology, 1*, 28. doi: 10.3389/fpsyg.2010.00028

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence-Teleoperators and Virtual Environments, 12*(5), 481–494. doi: 10.1162/105474603322761289

Nowak, K. L., Hamilton, M. A., & Hammond, C. C. (2009). The effect of image features on judgments of homophily, credibility, and intention to use as avatars in future interactions. *Media Psychology, 12*(1), 50–76. doi: 10.1080/15213260802669433

Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on Online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication, 11*(1), 153–178.

Park, S.-B., & Chung, N. (2011). Mediating roles of self-presentation desire in online game community commitment and trust behavior of massive multiplayer online role-playing games. *Computers in Human Behavior, 27*(6), 2372–2379. doi: 10.1016/j.chb.2011.07.016

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*, 437–442.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Exp Brain Res, 209*, 481–493. doi: 10.1007/s00221-011-2579-1

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience, 2*(1), 67–70. doi: 10.1093/scan/nsm006

Puce, A., Syngeniotis, A., Thompson, J. C., Abbott, D. F., Wheaton, K. J., & Castiello, U. (2003). The human temporal lobe integrates facial form and motion: Evidence from fMRI and ERP studies. *Neuroimage, 19*, 861–869. doi: S1053811903001897 [pii]

Reeves, B., & Nass, C. (1996). *The Media Equation: How people treat computers, television, and new media like real people and places*. Stanford, CA: CSLI.

Rossion, B. (2008). Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. *Neuroimage, 40*, 423–426. doi: S1053-8119(07)00990-1 [pii] 10.1016/j.neuroimage.2007.10.047

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *Neuroimage, 30*, 1088–1096.

Strafling, N., Fleischer, I., Polzer, C., Leutner, D., & Kramer, N. C. (2010). Teaching learning strategies with a pedagogical agent: The effects of a virtual tutor and its appearance on learning and motivation. *Journal of Media Psychology: Theories, Methods, and Applications, 22*(2), 73–83. doi: 10.1027/1864-1105/a000010

Taylor, L. D. (2011). Avatars and emotional engagement in asynchronous online communication. *Cyberpsychology Behavior and Social Networking, 14*(4), 207–212. doi: 10.1089/cyber.2010.0083

Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior, 27*(2), 741–749. doi: http://dx.doi.org/10.1016/j.chb.2010.10.018

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17*(1–3), 257–280.

Weber, R., Mangus, J. M., & Huskey, R. (2015). Brain imaging in communication research: A practical guide to understanding and evaluating fMRI studies. *Communication Methods and Measures, 9*(1), 1–26. doi: 10.1080/19312458.2014.999754

Weibel, D., Stricker, D., Wissmath, B., & Mast, F. W. (2010). How socially relevant visual characteristics of avatars influence impression formation. *Journal of Media Psychology: Theories, Methods, and Applications, 22*(1), 37–43.

Wiggett, A. J., & Downing, P. E. (2008). The face network: overextended? (Comment on: "Let's face it: It's a cortical network" by Alumit Ishai). *Neuroimage, 40*(2), 420–422. doi: S1053-8119(07)01103-2 [pii] 10.1016/j.neuroimage.2007.11.061

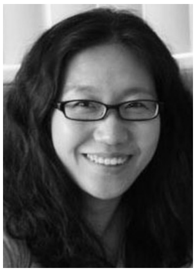
Date of acceptance: March 16, 2015

Robert F. Potter (PhD, Indiana University) is an associate professor of communication science in The Media School at Indiana University. His research interests include cognitive and emotional response to audio and psychophysiological methods.

Robert F. Potter
Department of Telecommunications
Indiana University
1229 E. Seventh St.
Bloomington, IN 47405
USA
Tel. +1 812 856-2546
E-mail rfpotter@indiana.edu

Thomas James received his PhD in neuroscience from Western University in London, Ontario, Canada, and pursued postdoctoral studies at Vanderbilt University in Nashville, Tennessee, USA. He is currently the head of the Perception and Neuroimaging Laboratory at Indiana University Bloomington, where he studies brain mechanisms for perceptual decision making.

Sungkyoung Lee (PhD, Indiana University) is an assistant professor in the school of journalism at University of Missouri, Columbia. Her research focuses on understanding of mediated message processing and designing effective health communication messages.



Sunah Kim received her double PhD in cognitive science and neural science from Indiana University, Bloomington, and pursued postdoctoral studies at the University of California, Berkeley, USA. She is currently conducting user experience research in the areas of crossmodal perception, human–computer interaction, and innovative health services at Samsung Electronics.

Ryan Stevenson (PhD, Indiana University) is currently a Banting Postdoctoral Fellow in the Department of Psychology at the University of Toronto. He is interested in audiovisual sensory processing and interactions between sensory perception and higher order cognitive functions. His work focuses on sensory perception in development and developmental disabilities, including autism.

Annie Lang (PhD, University of Wisconsin at Madison) is distinguished professor of telecommunications and cognitive science at Indiana University. She studies mediated message processing. She is a former editor of *Media Psychology,* a Fellow of the International Communication Association, and recipient of the International Communication Association (ICA) Steven H. Chaffee Career Achievement Award.