The uncanny valley does not interfere with level 1 visual perspective taking

Karl F. MacDorman*, Preethi Srinivas, Himalaya Patel

Indiana University School of Informatics, 535 West Michigan Street, Indianapolis, IN 46202, USA

1. Introduction

Anthropomorphism tends to increase with human realism; people are more likely to ascribe human qualities to objects that look and act human—and to evaluate them positively (Ho & MacDorman, 2010; Wang, Baker, Wagner, & Wakefield, 2007; Yee, Bailenson, & Rickertsen, 2007). However, Mori (1970/2012) noted a startling exception. As many of an object’s features are made to resemble those of a human being, vestigial nonhuman features become salient (Fig. 1). Mismatches among human and nonhuman features strike discords, and the overall effect is eerie (MacDorman, Green, Ho, & Koch, 2009; MacDorman & Ishiguro, 2006; Mitchell, Ho, Patel, & MacDorman, 2011; Seyama & Nagayama, 2007; Tinwell, Grimshaw, & Williams, 2011). This dip in the observer’s affinity for the object is called the uncanny valley (Mori, 1970/2012). The term is now commonly applied to characters in animated films and videogames to which viewers cannot relate because of flaws in how they were computer-modeled, animated, and rendered (Geller, 2008; Pollick, 2010). In particular, the combination of realistic characters and animation based on the motion capture of live actors is blamed for audiences’ difficulty in experiencing empathy for characters in films like Mars Needs Moms (2011), A Christmas Carol (2009), and The Polar Express (2004; Butler & Joschko, 2009; Freedman, 2012; Hale, 2011; Hodgins, Jörg, O’Sullivan, Park, & Mahler, 2010; Melina, 2011).

Empathy is the ability to project one’s self into the body and circumstances of another to make sense of that individual’s actions by experiencing congruent feelings, beliefs, intentions, motivations, and sensations (de Vignemont & Singer, 2006). Because the human capacity for empathy is essential for the success of a story (Coplan, 2006; Zillmann, 2006), suppression of empathy could easily ruin a film or video game. Thus, it is important not only to identify any phenomenon in computer animation that suppresses empathy but also to circumvent it through an understanding of its elicitors and underlying mechanisms. However, the effect of the uncanny valley on empathy has not been investigated systematically. Matters are complicated by the fact that empathy is not a unitary process but a collection of partially dissociable neurocognitive processes that may be grouped under the broad headings of cognitive, emotional, and motor empathy (Blair, 2005; Decety & Moriguchi, 2007). For example, the ability to take the perspective of others visually or cognitively or to imagine others in a hypothetical situation may be grouped under cognitive empathy or theory of mind (Davis, 1983; Frith, 2001; Leslie, 1987); the ability to experience the emotions of others “contagiously” or to sympathize with those in need and desire to help them may be grouped under emotional empathy (Hatfield, Cacioppo, & Rapson, 1993; Moriguchi et al., 2007; Nummenmaa, Hirvonen, Parkkola, & Hietanen, 2008); and the ability to mimic others nonconsciously or synchronize with their movements may be grouped under...
When the number of dots differed, performance declined, indicating that they could neither fully ignore the character's perspective nor fully ignore their own perspective when making judgments from the character's perspective. However, there has been little systematic exploration of the underlying causes of this interference.

One possible explanation is that the character's presence activated automatic perspective-taking processes in the brain that cannot be fully inhibited by conscious intent. The mirror neuron system is a candidate for these processes. Mirror neurons in the premotor cortex are active both when an individual is performing an intentional action and when the individual is observing someone else performing the same action (Gallese, 2003; Keysers & Gazzola, 2006). The theoretical assumption of mirror neuron research is that one is able to infer the intentions of others based on one's own intentions in a similar situation, because the same circuits in the brain are shared in taking the perspective of one's self or of others (Lombardo et al., 2010). Some neuroimaging studies have found characters with high human photorealism (i.e., those that visually appear more human) activate the mirror neuron system and other shared circuits more than characters with low human photorealism, though other studies have found no such effect (Gazzola, Rizzolatti, Wicker, & Keysers, 2007). Tai, Scherl, Brooks, Sawamoto, and Castiello (2004), for example, found a significant neural response in participants' left premotor cortex when watching an action performed by a human being but not when watching the same action performed by a robot. Krach et al. (2008) discovered a linear relation between a robot's human photorealism and cortical activation in the medial frontal cortex and the right temporal parietal junction. Kilner, Paulignan, and Blakemore (2003) found that observing a human making incongruent movement had a significant interference effect on participants, but that observing a robotic arm making incongruent movements did not. Their research aligns with Press (2011) who found that the “action observation network,” which translates observed actions into motor codes, is biologically tuned and not engaged by nonbiological agents. These findings suggest that a character's human photorealism may differentially activate brain processes that interfere with the ability to take one's own perspective.

In creating a theory to explain this effect, past research would lead us to expect that higher human photorealism in a character would enhance perspective taking from the character's perspective through increased activation of shared circuits. This increased activation could interfere with making self-perspective judgments, when the observer's perspective differs from the character's perspective. In addition, if the uncanny valley phenomenon suppresses cognitive empathy and, more specifically, level 1 visual perspective taking, an eerie character should activate shared circuits less than a character that is not eerie. Therefore, the effect of high eeriness on level 1 perspective taking should be similar to that of low human photorealism.

Thus, to explain how the uncanny valley phenomenon impairs perspective taking, a novel theory is proposed: The more human or less eerie a character looks, the more it interferes with level 1 visual perspective taking when the character's perspective differs from that of the human observer. This theory is tested experimentally by means of a dot-counting task in which the participant and character face either the same number of dots (congruent trials) or a different number of dots (incongruent trials), and the participant is instructed to report the number of dots visible either from the perspective of self or other (i.e., the participant's own perspective or the character's perspective).

1.1. Hypotheses

Consistent with the findings of Samson et al. (2010), the proposed model predicts higher performance in congruent trials—
is, a faster response time and fewer errors—for both the self and other perspectives (Hypothesis 1). The model also predicts that characters having higher photorealism inhibit performance in self–incongruent trials (Hypothesis 2a) and facilitate performance in the other three combinations of perspective and congruence: self–congruent (Hypothesis 2b), other–incongruent (Hypothesis 2c), and other–congruent (Hypothesis 2d). Finally, the model makes the same four predictions for characters with lower eeriness than for characters with higher eeriness: lower performance in self–incongruent trials (Hypothesis 3a) and higher performance in self–congruent (Hypothesis 3b), other–incongruent (Hypothesis 3c), and other–congruent trials (Hypothesis 3d).

2. General method

2.1. Overview

This study consists of four experiments in which participants perform on a computer a multiple-trial task that begins with the presentation of a character in a room. Experiments 1, 2, and 3 consist of a dot-counting task; Experiment A1, described in the Appendix, consists of a direction determination task performed as a control for the characters in Experiments 1 and 2.

Experiments 1 and 2 use three-dimensional computer-modeled characters that were divided into five classes based on similarity and that vary greatly in their human photorealism: inanimate objects, robots, fantasy beings, nonhuman animals, and human beings. The only difference between Experiments 1 and 2 is whether the perspective-taking instructions, self or other, are randomly interspersed or segregated into two blocks. Experiment 3 imposes stricter experimental controls over a narrower range of characters, because only male human characters are used based on three young White men (CVL Face Database; Solina, Peer, Batagelj, Juvan, & Kovac, 2003). Each man is presented in three ways to vary human photorealism: as a two-dimensional illustration, as a three-dimensional computer model, and as photographed. Nonhuman characters were included in Experiments 1 and 2 to search for broad trends in how anthropomorphic properties may affect perceived eeriness and interfere with level 1 visual perspective taking. Experiment 3 focuses on these dimensions more narrowly by depicting human characters at three levels of photorealism; the results have a direct bearing on the uncanny valley in computer animation.

2.2. Participants

A total of 895 participants, aged 18–66 (\(M_{\text{age}} = 23.6, SD_{\text{age}} = 7.3\)), were recruited by electronic mail from a randomized exhaustive list of undergraduates attending a nine-campus Midwestern public university. Experiment 1 was conducted in November 2010. Experiment 2, a variation of Experiment 1, was conducted in July 2011. Experiment 3 was conducted in May 2011. Experiment A1 was conducted in July 2011. Participants were randomly assigned to the four experiments. Between 200 and 240 unique individuals participated in each experiment. No individual participated in more than one experiment or had prior knowledge of the experiment.

2.3. Stimuli and procedure

Participants completed their assigned experiment through an access-controlled website. All experiments used a within-group design, in which the same stimuli were presented in the same trials; however, the order of trials was randomized for each participant. The stimuli filled a region of the computer screen 800 pixels wide by 600 pixels high. A red X indicated an incorrect response, and the trial continued until the participant made the correct response. There was a 1000-ms interval between trials. Each experiment took 4–8 min to complete. During each experiment, a pause screen appeared at four evenly spaced intervals, giving the participant a chance to rest. To decrease learning effects, each experiment was preceded by instructions and examples of correctly completed trials (Fig. 2). Each example included a character used in the corresponding experiment, a given perspective, a congruent or incongruent arrangement of dots between the self and other perspectives, and the correct number of dots visible from the given perspective. Participants were instructed to respond rapidly and accurately.

2.3.1. Dot-counting task

In the dot-counting task, a character was presented in the middle of a room, facing toward either the left wall or the right wall; on each of these two walls simultaneously appeared between zero and three dots; and after a 500-ms delay at the top of the scene appeared the instruction self or other. Self indicated the dots were to be counted from the participant’s own visual perspective; other indicated the dots were to be counted from the visual perspective of the character in the room.

There were 12 possible combinations of character direction (either left-facing or right-facing) and patterns of dots on the left and right walls (Table 1). Each configuration was presented once with the instruction self and once with the instruction other. Thus,
there were 24 trials per character, which divided equally into the congruent and incongruent perspective-taking conditions. For each trial, response times (RTs) were recorded as the elapsed time in milliseconds from the onset of the instruction until the registering of the correct response. Errors were defined as initial responses that were incorrect but otherwise valid (e.g., 0, 1, 2, or 3).

The design of the dot-counting task differed somewhat from that of Samson et al. (2010). In their experiment, the instruction was presented first, then a number, and then the room with the dots and one of the characters. For a given instruction, participants indicated whether the number matched the correct response instead of simply indicating the correct response. Their design was modified, because it primed participants on what to ignore by giving the instruction first. The current design presents the character in the room first to allow time for the character to activate brain processing before the instruction appears. This design is intended to enhance sensitivity in the self condition by preventing early filtering of the character through selective attention. Because matched and mismatched trials were unbalanced in the design of Samson et al. (2010), they scored only matched trials. However, because either 9 or 10 characters were used in each experiment in this study, to reduce fatigue effects all trials were scored. To reduce the number of trials further, the number of configurations of character direction and dot pattern were reduced from 20 to 12, and configurations were not repeated for the same character.

2.3.2. Animacy and eeriness ratings

Following the dot-counting task in Experiments 1, 2, and 3, participants rated the characters on semantic differential scales selected from indices designed to measure perceived animacy and eeriness (Ho & MacDorman, 2010). The animacy anchors were artificial–natural, inanimate–living, and without definite lifespan–mortal. The eeriness anchors were bland–creepy, boring–weird, uninspiring–spine-tingling, and numbing–freaky.

2.4. Data analysis

Data were analyzed using SPSS (version 19). Participants were removed using two methods: First, Chebyshev’s inequality and its empirical rule were applied, thus, removing participants with any trial having a z score > 3. Second, excluded from the analysis were data from participants who made one or more anticipatory or unattentive responses, operationalized as responses outside the range of 300–3000 ms (Greenwald, Nosek, & Banaji, 2003). Both impossibly fast responses (i.e., those initiated before stimulus perception) and very slow responses indicate nonconformance with the instructions.

The central tendency and dispersion of response times (RTs) and percentage of errors (PEs) are reported using the mean and standard deviation, respectively, to maintain consistency with prior literature. Statements of statistical inference are based on a significance level (α) of .05.

To correct positive skew in all four experiments, RT was transformed using one of two methods, depending on the corresponding statistical procedure: Before obtaining F and t statistics, which were used to test Hypothesis 1, each recorded value was replaced by its base-10 logarithm (log10). Before performing linear regression, which was used to test Hypotheses 2 and 3, each recorded value was replaced by the difference between the original value and the arithmetic mean for that participant from all trials having the present trial’s combination of perspective and congruence. By retaining the original unit of measurement (milliseconds elapsed), this relative-difference transformation simplified interpretation of the data. To preserve consistency, whenever the RT data were corrected using the relative-difference transformation, the PE data were corrected similarly.

3. Experiment 1

3.1. Participants

Experiment 1 had 240 participants (M_age = 22.5, SD_age = 6.8; 36% male).

3.2. Stimuli and procedure

Ten 3-dimensional characters of varying degrees of human likeness were rendered in 3D animation software (Autodesk Maya 2011): an arrow, a chair, Cramer (a humanoid robot with a boxy, bright red exterior), R2-D2 (a robot from the film series Star Wars), Creature (an elephantine being), a zombie, a bee, a bear, a man, and a woman (Fig. 3). The total number of trials was 240 (10 characters × 12 combinations of character direction and dots on walls × 2 instructions). In this experiment the order of the trials was fully randomized; thus, the instruction self was randomly interspersed with the instruction other.

3.3. Results and discussion

3.3.1. Ratings of animacy and eeriness

The mean ratings for animacy (Cronbach’s z = .90) and eeriness (z = .93) were calculated for the 10 characters from Experiments 1 and 2 (Fig. 4). The inanimate objects and robots were rated as low in animacy, the fantasy beings as moderate, and the nonhuman animals and humans as high. Inanimate objects were rated as low in eeriness, the humans, robots, and nonhuman animals were rated as moderate, and the fantasy beings were rated as high.
Task performance was particularly poor for three characters: R2-D2, the arrow, and the chair (Fig. 5). Because robots are more relevant than inanimate objects to the present research, R2-D2’s results were investigated further. To determine whether R2-D2’s mostly symmetrical form was confusing, a subsequent two-alternative forced-choice study was conducted in which participants rapidly identified the direction each character was facing (Appendix A). The direction determination task presented the same 10 characters used in Experiments 1 and 2. The results showed that there was not a significant learning effect and, thus, the results for R2-D2 from Experiment 1 were not caused by a lack of familiarity with the character or its orientation.

### 3.3.2. Inferential analysis

After the log_{10} transformation, the RT distribution became sufficiently normal (Kolmogorov–Smirnov $D = 0.061$, $p = .083$). A $2 \times 2$ repeated measures ANOVA was performed for the dependent variable RT (log_{10} transformed). Both perspective (the instruction self or other) and congruence (congruent or incongruent number of dots visible to the participant and character) were included as within-subjects factors. Post-hoc paired t tests were performed with the Bonferroni correction for multiple comparisons. McNemar’s test was performed for the dependent variable PE.

### 3.3.3. Congruence

Hypothesis 1, which stipulates higher performance in congruent trials, was supported by both the RT and PE data (Fig. 5). RT was significantly faster ($t(239) = 741.31$, $p < .001$, $r = .49$) in congruent trials ($M = 948$ ms) than in incongruent trials ($M = 1180$ ms), and PE was significantly lower ($\chi^2(1, N = 31,373) = 1518.85$, $p < .001$) in congruent trials ($M = 3.4\%$) than in incongruent trials ($M = 11.3\%$).

### 3.3.4. Human photorealism

The 10 characters were sorted into five levels of human photorealism from negligible to high: inanimate objects (arrow and chair), robots (Cramer and R2-D2), fantasy beings (Creature and zombie), nonhuman animals (bee and bear), and humans (man and woman). Both the RT and the PE distributions became sufficiently normal after performing the relative-difference transformation (Kolmogorov–Smirnov $D = 0.065$, $p = .075$; $D = 0.248$, $p = .061$). This transformation was used in testing Hypotheses 2 and 3.

Hypothesis 2 stipulates that performance increases (i.e., RTs and PEs decrease) as the level of human photorealism increases for all combinations of perspective and congruence except self–incongruent, for which performance decreases (i.e., RTs and PEs increase). Hypothesis 2 was tested using linear regression: In each of the four main treatment conditions, a line was fitted to the RT and PE data to compare the effect of increasing humanness with the predicted slope (Fig. 6).

For each of the four statements in Hypothesis 2, this method produced two indicators, one for each of the two dependent variables. These claims are examined in order: Hypothesis 2a, which predicts a positive slope in self–incongruent trials, was unsupported.
for both RT and PE ($B = -10.44, t[2399] = -3.77$, one-sided $p > .999$; $B = -0.004, t[2399] = -2.32, p = .990$). Hypothesis 2b, which predicts a negative slope in other–incongruent trials, was supported for RT ($B = -12.83, t[2399] = -4.67, p < .001$) with a small effect size ($\beta = -.095$) but was unsupported for PE ($B = -0.001, t[2399] = -0.49, p = .311$). Hypothesis 2c, which predicts a negative slope in self–congruent trials, was supported for RT ($B = -5.07, t[2399] = -2.57, p = .005$) with a small effect size ($\beta = -.052$) but was unsupported for PE ($B = 0.000, t[2399] = 0.11, p = .544$). Hypothesis 2d, which predicts a negative slope in other–congruent trials, was supported for RT ($B = -15.23, t[2399] = -6.85, p < .001$) with a small effect size ($\beta = -.139$) and supported for PE ($B = -0.003, t[2399] = -3.44, p < .001$) with a small effect size ($\beta = -.070$). Overall, the prediction that increasing human photorealism inhibits performance in self–incongruent trials and facilitates performance in all other trials was unsupported.

Fig. 5. Performance by character in Experiments 1–3. Although experimental controls on the effects of extraneous factors were the most stringent in Experiment 3, the experimental effects of the levels of human photorealism were the smallest. For this and all subsequent figures with error bars, the bars indicate the 95% confidence interval of the true mean.
3.3.5. Eeriness

To test Hypothesis 3, the 10 characters were sorted into three levels of eeriness based on participant ratings: Inanimate objects were grouped as low in eeriness, the robots, nonhuman animals, and humans as moderate, and the fantasy beings as high. As reported previously, both the RT and the PE distributions became sufficiently normal after performing the relative-difference transformation. Hypothesis 3 stipulates performance increases (i.e., RTs and PEs decrease) as the level of eeriness decreases for all combinations of perspective and congruence except self–congruent, for which performance decreases (i.e., RTs and PEs increase). To test Hypothesis 3 in each of the four main conditions, a regression line estimating the effect of eeriness was fitted to the RT and PE data (Fig. 7).

This method produced two indicators for each of the four statements in Hypothesis 3. These are examined in order: Hypothesis 3a, which predicts a negative slope in self–incongruent trials, was unsupported for RT ($B = -0.59, t(2399) = -1.55, \text{one-sided } p = .062$) but was supported for PE ($B = -0.008, t(2399) = -1.96, \text{one-sided } p = .026$) with a small effect size ($\beta = -.040$). Hypothesis 3b, which predicts a positive slope in other–incongruent trials, was unsupported for both RT and PE ($B = -30.29, t(2399) = -4.93, p > .999; B = 0.002, t(2399) = 0.51, p = .306$). Hypothesis 3c, which predicts a positive slope in self–congruent trials, was unsupported for both RT and PE ($B = 3.16, t(2399) = 0.71, p = .238; B = -0.001, t(2399) = -0.53, p = .702$). Hypothesis 3d, which predicts a positive slope in other–congruent trials, was unsupported for both RT and PE ($B = -24.02, t(2399) = -4.81, p > .999; B = -0.005, t(2399) = -2.05, p = .980$). Overall, the prediction that increasing eeriness facilitates performance in self–incongruent trials was weakly supported, and the prediction that increasing eeriness inhibits performance in all other trials was unsupported.

3.4. Experiment summary

The results of Experiment 1 support Hypothesis 1. In incongruent trials participants were influenced not only by what they saw when interpreting the characters’ visual perspective (egocentric intrusion in other trials) but also by what the characters saw when interpreting their own visual perspective (altercentric intrusion in self trials). Across congruence conditions RTs involving egocentric intrusions were slower than RTs involving altercentric intrusions. Although a similar effect was observed for PEs in incongruent trials, the effect was reversed in congruent trials. Thus, the overall results support neither Hypothesis 2 nor Hypothesis 3.

4. Experiment 2

An alternative explanation of the effect of congruence in Experiment 1 is that it may have been caused by cognitive load incurred by frequent intertrial changes in perspective. To investigate this alternative explanation, Experiment 2 presented trials using each perspective instruction word (self and other) in separate blocks.

4.1. Participants

Experiment 2 had 200 participants ($M_{\text{age}} = 23.4, SD_{\text{age}} = 6.6; 31\% \text{ male}$).4

4 Of the 215 participants completing all parts of the experiment, 15 were excluded from data analysis for making one or more responses outside the range of 300–3000 ms.
4.2. Stimuli and procedure

Experiment 2 was identical to Experiment 1 except the 240 trials were partitioned into two counterbalanced blocks: one for the instruction self and the other for the instruction other. Within those blocks the order of trials was randomized between participants.

4.3. Results and discussion

4.3.1. Performance by character

Blocking appeared to increase performance in this experiment (M_RT = 750 ms, M_PE = 3%) compared with Experiment 1 (M_RT = 1051 ms, M_PE = 7%; Fig. 5). However, performance remained relatively poor for R2-D2, the arrow, and the chair. As in Experiment 1, participants were slowest and made the most errors for R2-D2. Although all trials with the instruction other appeared in one block, RTs were slowest for the arrow and chair.

4.3.2. Inferential analysis

After the log_{10} transformation, the RT distribution became sufficiently normal (Kolmogorov–Smirnov D[200] = 0.068, p = .808). Data from Experiment 2 was analyzed using the same methods as Experiment 1.

4.3.3. Congruence

As in Experiment 1, Hypothesis 1 (higher performance in congruent trials) was supported by both the RT and PE data. RT was significantly faster (t[199] = 23.06, p < .001, r = .85) in congruent trials (M = 731 ms) than in incongruent trials (M = 785 ms), and PE was significantly lower (χ^2(1, N = 25,431) = 9.607, p = .002) in congruent trials (M = 3.2%) than in incongruent trials (M = 3.6%).

4.3.4. Human photorealism

As in Experiment 1, the 10 characters were sorted into five levels of human photorealism: inanimate objects (arrow and chair), robots (Cramer and R2-D2), fantasy beings (creature and zombie), nonhuman animals (bee and bear), and humans (man and woman). The relative-difference transformation, both the RT and PE distributions became sufficiently normal (Kolmogorov–Smirnov D[7960] = 0.065, p = .78; D[7960] = 0.336, p = .71). As stated previously, Hypothesis 2 stipulates that performance increases (i.e., RTs and PEs decrease) as the level of human photorealism increases for all combinations of perspective and congruence except self–incongruent, for which performance decreases (i.e., RTs and PEs increase). As in Experiment 1, Hypothesis 2 was tested using linear regression: In each of the four main treatment conditions, a line was estimated the effect of human photorealism was fitted to the RT data (Fig. 8).

This method produced two indicators for each of the four statements in Hypothesis 2, one for each of the two dependent variables. These are examined in order: Hypothesis 2a, which predicts a positive slope in self–incongruent trials, was supported for both RT and PE (B = −.211, t[1989] = −.16, one-sided p = .564; B = −.001, t[1989] = −.099, p = .839). Hypothesis 2b, which predicts a negative slope in other–incongruent trials, was supported for RT (B = −.766, t[1989] = −3.92, p < .001) with a small effect size (r = −.088) but was unsupported for PE (B = 0.000, t[1989] = −0.44, p = .331). Hypothesis 2c, which predicts a negative slope in self–congruent trials, was supported for RT (B = −2.571,
Fig. 8. Using linear regression, the effects of human photorealism in Experiment 2 were estimated on the two outcome variables, RT (individually adjusted) and PE, for each combination of congruence and perspective. The model predicts a positive slope for RT and PE in self–incongruent trials (Hypothesis 2a) and a negative slope in all other trials (Hypotheses 2b–d). Within each of these four main conditions, the level of human photorealism was at best a minor predictor of performance in the expected direction.

4.3.5. Eeriness

To test Hypothesis 3, the 10 characters were sorted into three levels of eeriness based on participant ratings: Inanimate objects were grouped as low in eeriness, the robots, nonhuman animals, and humans as moderate, and the fantasy beings as high. After the relative-difference transformation, both the RT and PE distributions became sufficiently normal (Kolmogorov–Smirnov D/7960 = 0.065, p = .078; D/7960 = 0.336, p = .071). Hypothesis 3 stipulates performance increases (i.e., RTs and PEs decrease) as the level of eeriness decreases for all combinations of perspective and congruence except self–incongruent, for which performance decreases (i.e., RTs and PEs increase). To test Hypothesis 3 in each of the four main conditions, a regression line estimating the effect of eeriness was fitted to the RT and PE data (Fig. 9).

This method produced two indicators for each of the four statements in Hypothesis 3. These are examined in order: Hypothesis 3a, which predicts a negative slope in self–incongruent trials, was unsupported for both RT and PE (β = -1.04, t[1989] = -0.36, one-sided p = .361; B = 0.001, t[1989] = 0.38, one-sided p = .647). Hypothesis 3b, which predicts a positive slope in other–incongruent trials, was unsupported for both RT and PE (β = -25.45, t[1989] = -8.31, p > .999; B = 0.003, t[1989] = 1.43, p = .076). Hypothesis 3c, which predicts a positive slope in self–congruent trials, was unsupported for both RT and PE (β = -3.17, t[1989] = -0.95, p = .303). Hypothesis 3d, which predicts a positive slope in other–congruent trials, was unsupported for both RT and PE (β = -21.50, t[1989] = -5.58, p > .999; B = -0.002, t[2399] = -0.71, p = .760). Overall, the predictions that increasing eeriness facilitates performance in self–incongruent trials and inhibits performance in all other trials were unsupported.

4.4. Experiment summary

The blocking introduced in this experiment decreased both RTs and PEs compared with Experiment 1. However, blocking eliminated neither egocentric nor altercentric intrusions. As in Experiment 1, across congruence conditions RTs involving egocentric intrusions were greater than RTs involving altercentric intrusions. The two intrusion effects produced similar PEs. Thus, although the participants were explicitly instructed to take a specific perspective, and the perspective within each block was unchanged, the participants were unable to ignore the irrelevant perspective. However, as in Experiment 1 and contrary to predictions, neither human photorealism nor eeriness mediated the congruence effect. Thus, the results of Experiment 2 support Hypothesis 1 but not Hypothesis 2 or 3.

5. Experiment 3

Experiments 1 and 2 were created to determine whether increasing the level of human photorealism, or decreasing the level of eeriness, inhibited performance during self–incongruent trials
and facilitated performance during all other trials. However, the previous experiments were exploratory in nature insofar as they used characters varying widely in type and appearance; thus, extraneous factors could have influenced accuracy and response times in the dot-counting task instead of the characters’ human photorealism and eeriness. This experiment imposed greater experimental control by using the same models at three levels of human photorealism. Models of the same gender, age, and race were used to limit further extraneous sources of variation.

5.1. Participants

Experiment 3 had 230 participants (M \( \text{age} = 25.3, SD_{\text{age}} = 8.1 \); 37% male).\(^5\)

5.2. Stimuli and procedure

The experimental paradigm was identical to Experiment 1 except for the characters used. Three young White men were presented at three levels of photorealism: two-dimensional illustration, three-dimensional computer model, and actual photograph. Hence, nine heads were used in this experiment. The faces of computer-modeled characters were made using FaceGen, which automatically created three-dimensional human head models from frontal and profile photographs (Fig. 10). Placing the nine heads into the dot-counting task resulted in 216 trials (3 models \( \times \) 3 levels of realism \( \times \) 12 combinations of character direction and dots on walls \( \times \) 2 instruction words).

5.3. Results and discussion

5.3.1. Ratings of animacy and eeriness

The mean ratings for animacy (Cronbach’s \( \alpha = .97 \)) and eeriness (\( \alpha = .91 \)) were calculated for the nine characters (Fig. 11). Based on participant ratings, the two-dimensional illustrations were grouped as low in animacy, the three-dimensional computer models as moderate, and the photographs as high.

5.3.2. Performance by character

Participants were slowest and made the most errors for the three-dimensional computer model and the two-dimensional illustrations of the third character. Participants were fastest and made the fewest errors for the photograph and the two-dimensional illustrations for the second character (Fig. 10).

5.3.3. Inferential analysis

After the log\(_{10}\) transformation, the RT distribution became sufficiently normal (Kolmogorov–Smirnov \( D[230] = .025, p = .200 \)). The dependent variables, RT (log\(_{10}\) transformed) and PE, were analyzed as in Experiments 1 and 2, after confirming homoscedasticity among groups. Both perspective (the instruction self or other) and congruence (congruent or incongruent number of dots visible to the participant and character) were included as within-subjects independent variables.

5.3.4. Congruence

Hypothesis 1, which stipulates higher performance in congruent trials, was supported by both the RT and PE data (Fig. 5). RT was sig-

---

\(^5\) Of the 243 participants completing all parts of the experiment, 13 were excluded from data analysis for making one or more responses outside the range of 300–3000 ms.
5.3.5. Human photorealism

The nine characters were sorted into three levels of human photorealism from low to high: 2D (two-dimensional illustrations of the three characters), 3D (three-dimensional computer models of the three characters), and PH (photographs of the three characters). After the relative-difference transformation, both the RT and PE distributions became sufficiently normal (Kolmogorov–Smirnov $D[230] = 0.067, p = .185; D[7956] = 0.251, p = .105$). As in Experiments 1 and 2, Hypothesis 2 (performance increases as the level of human photorealism increases for all combinations of perspective and congruence except self–incongruent, for which performance decreases) was tested using linear regression: In each of the four main treatment conditions, a line estimating the effect of human photorealism was fitted to the RT and PE data (Fig. 12).

This method produced two indicators for each of the four statements in Hypothesis 2 (Fig. 12). These are examined in order: Hypothesis 2a, which predicts a positive slope in self–incongruent trials, was unsupported for both RT and PE ($B = 1.235, t[1988] = 0.23$, one-sided $p = .411; B = 0.003, t[1988] = 0.79, p = .216$). Hypothesis 2b, which predicts a negative slope in other–incongruent trials, was unsupported for both RT and PE ($B = -0.844, t[1988] = -0.15, p = .441; B = 0.004, t[1988] = 1.55, p = .940$). Hypothesis 2c, which predicts a negative slope in self–congruent trials, was unsupported for RT ($B = 5.553, t[1988] = 1.41, p = .921$) but was supported for PE ($B = -0.003, t[1988] = -1.99, p = .023$) with a small effect size ($\beta = -.045$). Hypothesis 2d, which predicts a negative slope in other–congruent trials, was supported for RT ($B = -9.129, t[1988] = -2.14, p = .016$) with a small effect size ($\beta = -.048$) but was unsupported for PE ($B = -0.002, t[1988] = -0.95, p = .172$).

Overall, the prediction that increasing human photorealism inhibits performance in self–incongruent trials and facilitates performance in all other trials was unsupported.

5.3.6. Eeriness

To test Hypothesis 3, the nine characters were sorted into three levels of eeriness based on participant ratings: two-dimensional illustrations were grouped as low in eeriness, photographs of the three characters as moderate, and three-dimensional computer models as high. After the relative-difference transformation, both the RT and PE distributions became sufficiently normal (Kolmogorov–Smirnov $D[230] = 0.067, p = .185; D[7956] = 0.251, p = .105$). As in Experiments 1 and 2, to test Hypothesis 3 in each of the four main conditions, a regression line estimating the effect of human photorealism was fitted to the RT and PE data (Fig. 13).

This method produced two indicators for each of the four statements in Hypothesis 3. These are examined in order: Hypothesis 3a, which predicts a negative slope in self–incongruent trials, was unsupported for both RT and PE ($B = 10.398, t[1988] = 1.90$, one-sided $p = .971; B = 0.006, t[1988] = 1.99, p = .976$). Hypothesis 3b, which predicts a positive slope in other–incongruent trials, was unsupported for both RT and PE ($B = 1.716, t[1988] = 0.31, p = .380; B = -0.004, t[1988] = -1.29, p = .902$). Hypothesis 3c, which predicts a positive slope in self–congruent trials, was supported for RT ($B = 7.598, t[1988] = 1.94, p = .027$) with a small effect size ($\beta = .043$) but was unsupported for PE ($B = -0.003, t[1988] = -1.44, p = .925$). Hypothesis 3d, which predicts a positive slope in other–congruent trials, was unsupported for both RT and PE ($B = -1.650, t[1988] = -0.39, p = .651; B = 0.000, t[1988] = -0.12, p = .548$). Overall, the prediction that increasing eeriness facilitates performance in self–incongruent trials and inhibits performance in all other trials was unsupported.

5.4. Experiment summary

The results of Experiment 3 supported Hypothesis 1: performance (indicated by RT and PE) was worse in incongruent trials. As in Experiments 1 and 2, both egocentric and altercentric intrusions were observed. Although the egocentric intrusions were greater than the altercentric intrusions for RT, this trend was absent for PE. While PEs increased in self–incongruent trials, PE did not increase monotonically with the human photorealism of the characters. These results do not support Hypothesis 2 or 3.
6. General discussion

In the preceding experiments involving a visual perspective-taking task, it was predicted that two aspects of a character’s appearance, human photorealism and eeriness, moderate both allocentric and altercentric intrusions. This would provide evidence that the uncanny valley phenomenon explains part of the variance found by Samson et al. (2010). Although both kinds of intrusions were observed, the moderating effects were either undetectable or simultaneously small and inconsistent. In other words, neither human photorealism nor eeriness influenced performance appreciably. For example, the eeriest characters (e.g., Zombie, Creature, computer model of male 3) did not interfere the least with self-perspective performance, nor did the most human characters (e.g., Man, Woman, photograph of male 1, 2, and 3) interfere the most. The results support the intrusion effects found by Samson et al. (2010) but fail to support the present study’s predictions about human photorealism and eeriness. Other aspects of each character’s appearance may be more predictive of task performance.

Fig. 12. Using linear regression, the effects of human photorealism in Experiment 3 were estimated on the two outcome variables, RT (individually adjusted) and PE, for each combination of congruence and perspective. The model predicts a positive slope for RT and PE in self–incongruent trials (Hypothesis 2a) and a negative slope in all other trials (Hypotheses 2b–d). Within each of these four main conditions, the level of human photorealism was not a significant predictor of performance in the expected direction.

6. Acknowledgments

We would like to thank Chin-Chang Ho, Wade J. Mitchell, and Amy Shirong Lu for their extensive and helpful advice in revising this manuscript. We are grateful to Matthias Scheutz for providing the Cramer model, Chin-Chang Ho for preparing stimuli, and Karl Lee Loftis for implementing the application used in the experiments. This study has been approved by the IUPUI/Clarian Research Compliance Administration (IRB Study No. 1010002308).

Appendix A. Experiment A1

The dot-counting task in Experiments 1 and 2 presented a heterogeneous group of characters, which included objects with no eyes...
(arrow and chair), fantasy beings (Creature and zombie), and robots (Cramer and R2-D2). Some of the characters were novel and thus unfamiliar (e.g., Creature and Cramer). To determine the number of dots a character is facing when the instruction other is given, a participant typically must determine whether the character is facing toward the left or toward the right. Therefore, this experiment measures participants’ response times and percentage of errors in determining the direction of the characters from Experiments 1 and 2. The results can diagnose whether the visual system took longer to determine the direction of a particular character or misperceived the character’s direction, which would appear as a learning effect as performance improved with task feedback. Therefore, participants were given a direction determination task to aid with the interpretation of the results of Experiments 1 and 2.

A.1. Participants

Experiment A1 had 225 participants ($M_{\text{age}} = 23.5$, $SD_{\text{age}} = 7.0$; 40% male).\(^6\)

A.2. Stimuli and procedure

The direction determination task presented the same 10 characters used in Experiments 1 and 2. Participants indicated a character is facing when the instruction other is given, a participant typically must determine whether the character is facing toward the left or toward the right. Therefore, this experiment measures participants’ response times and percentage of errors in determining the direction of the characters from Experiments 1 and 2. The results can diagnose whether the visual system took longer to determine the direction of a particular character or misperceived the character’s direction, which would appear as a learning effect as performance improved with task feedback. Therefore, participants were given a direction determination task to aid with the interpretation of the results of Experiments 1 and 2.

A.3. Data analysis

Response times and percentage of errors were analyzed using descriptive and inferential statistics. One-way ANOVAs were used to compare character groups, and two-tailed paired samples $t$ tests were used to test for learning and fatigue effects.

A.4. Results

In the character direction task, the average RT across all characters and trials was 547 ms ($SD = 121$ ms), and the average PE was 3.8% ($SD = 19.1$%). Participants were slowest and made most errors for R2-D2 ($M_{\text{RT}} = 630$ ms, $SD_{\text{RT}} = 163$ ms; $M_{\text{PE}} = 7.4$, $SD_{\text{PE}} = 9.0$) and Cramer ($M_{\text{RT}} = 576$ ms, $SD_{\text{RT}} = 105$ ms; $M_{\text{PE}} = 6.8$, $SD_{\text{PE}} = 7.1$) and fastest and made fewest errors for the arrow ($M_{\text{RT}} = 504$ ms, $SD_{\text{RT}} = 96$ ms; $M_{\text{PE}} = 1.5$, $SD_{\text{PE}} = 3.0$; Fig. A1).

Based on one-way ANOVAs, after confirming homoscedasticity among groups, participants were significantly slower in performing the character direction task for the robot characters ($M_{\text{PE}} = 2.9$, $SE_{\text{PE}} = 0.4$) than the other characters ($M_{\text{PE}} = 7.1$, $SE_{\text{PE}} = 0.4$) and Cramer, than the other characters ($M_{\text{PE}} = 7.1$, $SE_{\text{PE}} = 0.4$). F(1, 2168) = 148.36, $p < .001$, with a medium effect size ($\eta^2 = 0.06$). Moreover, participants made significantly more errors for the arrow ($M_{\text{PE}} = 7.1$, $SE_{\text{PE}} = 0.4$) than the other characters ($M_{\text{PE}} = 2.9$, $SE_{\text{PE}} = 0.1$), F(1, 2168) = 200.6, $p < .001$, with a medium effect size ($\eta^2 = 0.08$).

The average RT on the first, second, and last (12th) trial was 565 ms ($SD = 279$ ms), 574 ms ($SD = 349$ ms), and 576 ms ($SD = 505$ ms), respectively. The mean PE on the first, second, and last trial was 4.0% ($SD = 19.7$, $SE = 3.1$), 4.1% ($SD = 19.7$, $SE = 3.1$), and 3.5% ($SD = 18.3$, $SE = 2.9$), respectively. The paired samples $t$ test was applied to determine whether these differences

---

\(^6\) Of the 289 participants who completed all parts of the experiment, 64 were excluded from data analysis for making one or more responses outside the range of 300–3000 ms.
were significant. Although participants responded 9 ms faster on the first trial ($M_{\log RT} = 2.727$, $SE_{\log RT} = 0.002$) than the second trial ($M_{\log RT} = 2.730$, $SE_{\log RT} = 0.002$), the difference was nonsignificant: $t(4148) = -1.257$, $p = .209$, two-tailed. Although participants responded 11 ms faster on the first trial than the last trial ($M_{\log RT} = 2.729$, $SE_{\log RT} = 0.002$), the difference was nonsignificant: $t(4132) = -.914$, $p = .361$, two-tailed. Moreover, the difference in PE between the first and second trial was nonsignificant: $t(4148) = .028$, $p = .978$, two-tailed; and the difference between the first and last trial was also nonsignificant: $t(4132) = -1.268$, $p = .205$, two-tailed.

To ascertain whether the slow RTs and greater PE for R2-D2 relative to the other characters was caused by an initial inability to perceive the character’s direction (e.g., owing to a lack of familiarity with the robot), the first trial was compared with the second and last trial to test for learning effects. However, there was no indication of learning effects, because RT and PE in fact increased from the first trial ($M_{RT} = 679$ ms, $SD_{RT} = 449$ ms, $MPE = 6.95$%,$SDPE = 25.4%$) to the second trial ($M_{RT} = 705$ ms, $SD_{RT} = 671$ ms, $MPE = 7.95$%, $SDPE = 27.0%$) and the last trial ($M_{RT} = 729$ ms, $SD_{RT} = 901$ ms, $MPE = 7.55$, $SDPE = 26.4%$). However, neither the increase in RT from the first trial ($M_{\log RT} = 2.789$, $SE_{\log RT} = 0.008$) to the second trial ($M_{\log RT} = 2.793$, $SE_{\log RT} = 0.008$), $t(417) = -.389$, $p = .698$, two-tailed, nor from the first trial to the last trial ($M_{\log RT} = 2.790$, $SE_{\log RT} = 0.009$), $t(412) = -0.134$, $p = .893$, was significant. Likewise, neither the increase in PE from the first trial to the second trial, $t(417) = .534$, $p = .594$, two-tailed, nor from the first trial to the last trial, $t(412) = 0.428$, $p = .669$, two-tailed, was significant.

A.5. Discussion

The results indicate that among the characters the visual system was significantly slower and less accurate in determining the direction of the robots, especially R2-D2. These differences could not be attributed to an initial failure to determine R2-D2’s direction, because there were no learning effects. The visual system was fastest in determining the direction of the arrow.

References


