Citation: Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces, *Computers in Human Behavior*, 29(3), 759-771, doi: 10.1016/j.chb.2012.11.021.

Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces.

Tyler J. Burleigh, Jordan R. Schoenherr, and Guy L. Lacroix

Department of Psychology, University of Guelph, 50 Stone Road East, Guelph, ON, Canada N1G 2W1
Department of Psychology, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6

***Abstract* — The uncanny valley theory (UVT) (Mori, 1970) proposes that when stimuli are defined by a near-perfect resemblance to humans they cause people to experience greater negative affect relative to when they have perfect human likeness (HL) or little to no HL. Empirical research to support this non-linear relationship between negative affect and HL has been inconclusive, however, and a satisfactory causal explanation has not yet emerged to explain existing findings. In two studies, we examined the relationship between HL and eeriness using digital human faces. First, we examined the relationship between HL and eeriness while controlling for extraneous variation in stimulus appearance. We created two HL continua by manipulating the facial proportions and polygon count of several digital human models. Second, we proposed and tested two causal hypotheses regarding the uncanny valley phenomenon that we refer to as category conflict and feature atypicality. We created two additional HL continua by manipulating the skin coloration and category membership of models. Across these continua we introduced an atypical feature. Our results suggest that HL is linearly related to emotional response, except under conditions where HL varies by category membership, suggesting that previous empirical findings might be explained as a category conflict.**

***Keywords* — human likeness; eeriness; uncanny valley; facial perception; categorization; social cognition**

1. THE UNCANNY VALLEY THEORY

Horror movies deliberately play on scary themes. They portray danger, disease, and violence in order to arouse and excite their audience. Indeed these aspects can be considered the defining characteristics of the horror genre (Walters, 2004). When people watch horror movies, they expect to be shocked, disgusted, and frightened. The extent to which these experiences are realized likely determines these films' success. Naturally, no one would expect a holiday children's movie to generate such reactions. It is particularly interesting then when such a movie, *The Polar Express* (Zemeckis, 2004), can be cited as an example of incidental horror (Geller, 2008; Kloc, 2009; Levy, 2004; Loder, 2004). In this film, animated digital replicas of real human actors were used to attain a high degree of realism. This animation process consisted of capturing the movement of a live actor and using this movement to direct a digital human. The result has been hailed by critics as "unnerving" (Hanel, 2008), "creepily unlifelike" (Keegan, 2009, p. 235), and as having a "bizarre wax-museum quality" (Levy, 2004, p. 1). Multiple authors have suggested that the near-perfect human likeness of the animation was responsible for people's negative reaction to it (e.g., Bartneck, Kanda, Ishiguro, & Hagita, 2007; Chaminade, Hodgin, & Kawato, 2007; MacDorman, Green, Ho, & Koch, 2009; Pollick, 2009). This phenomenon is known as the uncanny valley.

The uncanny valley theory describes how emotional reactions vary with perceived human likeness. Mori (1970) originally formulated the theory to describe people's reactions to robots, but it can be applied to anything possessing human-like qualities including digital animation (e.g., MacDorman et al., 2009) and voices (Tinwell & Grimshaw, 2009).

Figure 1 illustrates the theorized relationship between human likeness and emotional response. As the human likeness of a stimulus increases, an individual's emotional response to the stimulus becomes more positive, but when human likeness nears perfection, the individual's emotional response sharply declines and becomes strongly negative. The region immediately following this decline is the uncanny valley. Consequently, the uncanny valley theory makes the assumption that negative emotions are the result of a

stimulus's placement on a continuum of human likeness. Mori's (1970) uncanny valley theory does not, however, offer a causal explanation of eeriness; it does not explain why stimuli that are nearly perfectly human are perceived as eerie.

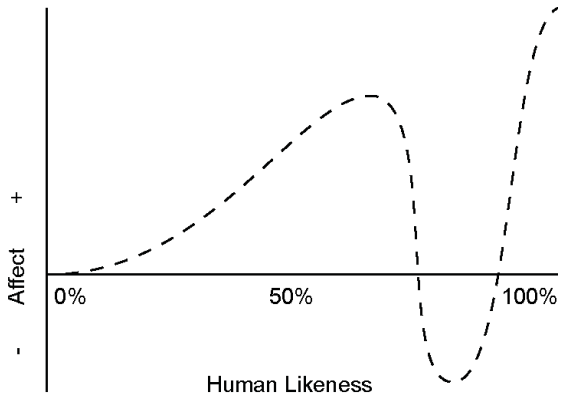

Fig. 1. The Uncanny Valley Function, as Proposed by Mori (1970).

According to MacDorman, the uncanny valley theory has already attained a dogmatic status (p. 2, MacDorman, as cited in Kloc, 2009). Consider, for example, the development of the movie *Avatar* (Cameron, 2009). When filmmaker James Cameron suggested using a motion capture technique like the one used in *The Polar Express* (Zemeckis, 2004), the graphics supervisors were reportedly "very uncomfortable, and they were fearful of it wrecking the company" (p. 236, as cited in Keegan, 2009). Mori (1970) had urged designers to avoid the negative consequences implied by his theory by deliberately sacrificing realism. If Cameron had reneged on motion capture, would *Avatar* have still been a record-breaking success? Certainly, Mori's (1970) advice is sound only if human likeness carries an inherent risk of eeriness.

The present experiments sought to critically re-examine the relationship between human likeness and eeriness as it relates to the uncanny valley. We will first review the extant literature, examining closely how human likeness has been operationalized in terms of subjective reports and stimulus generation. We will then test two novel hypotheses that might offer alternative explanations for the uncanny valley phenomenon as observed in some of these previous studies. Specifically, we hypothesize that an uncanny valley might emerge when human likeness is operationalized as a merger of human and non-human categories, or when a human likeness continuum is paired with one or more atypical features. Our secondary goal will be to clarify the negative emotions individuals experience in response to stimuli located in the "valley," an emotion that has been referred to as eeriness.

1.1. Previous research on the uncanny valley

In this section, the evidence that has been presented in support of the uncanny valley theory will be called into question. MacDorman and Ishiguro's (2006) experiments were the first to empirically support the uncanny valley theory's assumption that negative emotions are caused by stimulus placement on a continuum of human likeness. It will be argued that their findings are open to alternate interpretations and we will highlight similar studies (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2011; Seyama & Nagayama, 2007) that might also be challenged by the same difficulty. In contrast to these studies, MacDorman et al.'s (2009) paradigm did not find support for the uncanny valley theory. Limitations of their methodology which might have been the cause of null findings will be discussed.

One common approach that has been used to operationalize human likeness involves applying morphing software to human and robot images. Under these conditions, the human represents maximal human likeness and the robot represents minimal human likeness. Sometimes, as was the case with MacDorman and Ishiguro's (2006) study, a mid-point is also defined using an image of an android. MacDorman and Ishiguro (2006) generated a continuum consisting of 12 levels. These images were then presented and subjective reports were collected, including ratings of human likeness and emotional response. The human likeness scale was presented with anchor points of *mechanical* and *human-like*. Emotional responses were obtained using eeriness and familiarity scales. When subjective human likeness ratings were plotted across levels of the stimulus continuum, ratings were found to increase monotonically. However, eeriness and familiarity were non-linear, observing high and low peaks, respectively, approximately half-way along the continuum. Because these trends resembled the assumed relationship in Figure 1, the authors concluded that this result supported the uncanny valley theory. It is our position that MacDorman and Ishiguro's (2006) findings do not support a strict interpretation of the uncanny valley theory, however, as the non-linear trends might not have been caused by stimulus placement on a continuum of human likeness. Instead, we believe that these findings could be explained by two alternative accounts that we

will refer to as the *atypical feature hypothesis* and the *category conflict hypothesis*.

The *atypical feature hypothesis* states that the eeriness of a stimulus might be accounted for as a function of human likeness in combination with the presence of one or more atypical features. An examination of MacDorman and Ishiguro's (2006) stimulus generation procedure reveals that certain stimuli along the human likeness continuum, particularly those which were most eerie, were in possession of atypical features that were inherited from neighboring images in the morphing sequence. This can be seen in one morphing sequence where a large black dot appears on the centre of the robot's forehead. When this robot's image was entered into the morph, the black dot was preserved. Although this might be a feature that is normally diagnostic of a robot, the feature could be perceived in a manner that could cause negative affect when seen on a human (e.g., bullet hole, tumour, third eye...). Our interpretation is supported by the fact that the locations along the continuum where this feature appeared were also those which were maximally eerie and minimally familiar. However, stimuli located at these levels also possessed features that could not plausibly belong to a human, such as the robot's bulky metallic chassis. *Plausibility* is an important condition of our hypothesis which is satisfied if and only if the feature belongs to the same ontological category as the stimulus. In one of Seyama and Nagayama's (2007) experiments, an uncanny valley function was observed when an atypical human feature (enlarged eyes) was introduced across a continuum of human likeness. Thus, features that are unusual and also plausible should be expected to elicit negative affect as a function of human likeness.

From an evolutionary perspective, the atypical feature hypothesis may be understood as a signal detection problem (Nesse, 2005). In short, humans have evolved psychological mechanisms that enable the categorization of stimuli into fitness-relevant categories (e.g., rival, kin, mate, predator, or disease-carrier) using imperfect stimulus cues. These mechanisms have been calibrated to minimize false-positive errors (e.g., classifying a dangerous target as benign) at the expense of increasing false-negative errors (e.g., classifying a benign target as dangerous). Evolution erred on the side of caution to favor survival and reproduction (Nesse, 2005). If a feature on a conspecific stimulus is sufficiently atypical, then it can be expected to trigger one of these mechanisms independently of any real danger. If that feature were diagnostic of possible infection, then an individual would experience disgust (Park, Faulkner, & Schaller, 2003; Thornhill & Gangestad, 1993); if that feature were diagnostic of a physical threat, then an individual would experience fear (Buss, 2005; Cosmides & Tooby, 2000); if that feature were diagnostic poor mate quality, then an individual would experience negative attraction or disliking (Rhodes, 2006). Therefore, in terms of a negative response to a human-like stimulus with atypical features, at least three evolved psychological mechanisms could be involved, namely those corresponding to the emotions of fear, disgust, and disliking. Interestingly, Ho, Pramono, and Macdorman, (2008) found that these three emotions uniquely accounted for the variance in eeriness, among a variety of emotional responses to human-like stimuli with atypical features. However, given that Ho et al. (2008) used human-like robots as stimuli, the atypical features present were not necessarily plausible.

The *category conflict hypothesis* states that when human likeness is operationalized as a merger of human and non-human categories, stimuli which lie approximately mid-way between such categories will be perceived as ambiguous and thus elicit negative affect. MacDorman and Ishiguro's (2006) findings might also be interpreted in line with this account. Although their intention was to produce stimuli which varied in human likeness, the stimuli were generated by merging an exemplar from human and robot categories. Similarly, Seyama and Nagayama's (2007) merged human and doll categories. If a stimulus contains features that are highly diagnostic of both human and non-human categories, then we propose that this might have resulted in two mutually exclusive and conflicting stimulus interpretations. This hypothesis is reinforced by Botvinick et al.'s (2001) account of conflict, wherein conflict is defined as an instance of information processing in which multiple processing pathways share common resources and when each produces output that is incompatible with the other. This hypothesis might be further understood by analogy to classic cognitive dissonance theory (Festinger, 1957), which describes an internal state that occurs when an individual holds two incompatible ideas simultaneously and results in an emotional state of discomfort and unease (Elliot & Devine, 1994). Some have speculated that the uncanny valley might be caused by a mismatch between the human-likeness of stimulus features (e.g., Ho, Pramono, & Macdorman, 2008). In line with the category conflict account, we would argue that the negative affect in these instances is not due to some features being higher in human-likeness than others, but rather due to some

features being human and others being non-human, which results in competing stimulus interpretations.

Two recent studies could also be interpreted as providing support for such a category conflict account. First, Mitchell et al. (2011) presented participants with stimuli belonging to the categories of robot and human and measured their emotional response. Specifically, human and robot characters were presented uttering a phrase in a human or robot voice. They examined the impact of merging these categories such that a human would speak with a robot voice and a robot would speak with a human voice. Consistent with a category conflict account, eeriness increased when these categories were mixed as compared to when they were not. Secondly, Saygin et al. (2011) examined brain activity while presenting videos of a human, an android, or a robot performing basic motor tasks. They observed increased brain activity primarily in the parietal cortex when participants were viewing an android, as compared to a human or a robot. They suggest that this extra activity was due to prediction error arising from individuals using an incorrect neural model to predict the behaviour of the agent. For example, when an agent appeared human-like, a human neural model was activated to predict the agent's movement. Similarly, when an object appeared robot-like, a robot neural model was activated. Thus, prediction error occurred when an object appeared human but moved like a robot. However, this interpretation rests on the assumption that individuals relied on a single neural model (human or robot), but it might also be possible that androids activated both human and robot neural models, and that increased brain activity reflected a conflict between them. However, this possibility cannot be adequately judged, as conflict resolution processes are presumed to reside in the anterior cingulate cortex (Botvinick et al., 2001), and measurements of activity in this area were not reported. Thus, although the conflict account might be compatible with Saygin et al.'s (2011) findings, definite conclusions cannot be drawn from their data.

Whereas previous research might have, owing to methodological concerns, generated mixed results that support alternative accounts to the uncanny valley theory, there has been one study that addressed these concerns using an innovative stimulus generation procedure. MacDorman et al. (2009) used digital human models and generated a continuum of human likeness by modifying their texture and geometric realism. In two experiments, participants rated these stimuli on human likeness and eeriness scales. As expected, subjective human likeness increased in a linear fashion with increasing levels of realism. Contrary to the predictions of the uncanny valley theory, however, there was also a linear relationship between human likeness and eeriness. In an additional experiment, the eyes of models were enlarged across levels of human likeness in order to examine the effect of abnormalities on affective response. The authors found that the enlarged eyes were more eerie at higher levels of eye realism, a finding that supports the atypical feature hypothesis. One shortcoming of this study, however, was that stimuli were derived from a single digital human model. Thus, it is possible that properties unique to this model were responsible for the relationship observed between human likeness and emotional response, and for reactions to the atypical feature. Therefore, it is problematic to generalize from these findings. Nevertheless, MacDorman et al.'s (2009) methodology allows researchers to manipulate human likeness in systematic fashion that does not introduce potential confounds; that is, modifications that are not accompanied by features that are categorically ambiguous or that simply do not belong on a human face. Thus, an extension of this paradigm and a replication of these findings would appear to be the next logical step in examining the uncanny valley theory. In Experiment 1, we will build on this methodology and extend both the realism and the variety of models.

2. Present research approach

The present research was motivated to test the uncanny valley theory's assumption that the placement of stimuli on a continuum of human likeness is responsible for the uncanny valley phenomenon. Contrary to Mori's (1970) original hypothesis, however, we believe that human likeness continua, when generated with sufficient control over extraneous factors like atypical features and category membership, will produce linear patterns of emotional response. MacDorman et al.'s (2009) methodology appears particularly suited to pursue this inquiry. Thus, Experiment 1 will emulate their stimulus generation procedure. In this experiment, we also sought to explore the phenomenology of eeriness by examining its relationship with fear, disgust, and attractiveness. Additionally, we sought to obtain evidence in support of two alternative accounts of the uncanny valley phenomenon, namely the atypical feature and category conflict hypotheses. Thus, in Experiment 2 we generated two continua of human likeness. One continuum was limited to the ontological category of humans, and along

this continuum we deliberately introduced an atypical human feature. A second continuum was generated by merging human and non-human categories.

2.1. Research questions

To summarize, our hypotheses are as follows:

H1. When stimuli on human likeness continua do not include unusual human features or non-human features, then a linear relationship will be observed between human likeness and emotional response.

H2. Eeriness will generate positive relationships with fear, disgust, and a negative relationship with attractiveness.

H3. When human likeness is manipulated by merging the features of two different ontological categories, the uncanny valley phenomenon will be observed as a non-linear pattern of emotional response.

H4a. An atypical human feature will elicit greater eeriness when placed on a stimulus that is high in human likeness, than when placed on a stimulus that is low in human likeness.

H4b. Feature atypicality will additively interact with human likeness in predicting eeriness, such that the combined effect of human likeness and feature atypicality will elicit more eeriness than either effect alone. If support for H4a is found, then the impact of feature atypicality on eeriness is expected to increase as human likeness increases. If support for H4a is not found, then the impact of feature atypicality on eeriness is expected to increase as human likeness decreases.

3. EXPERIMENT 1

Our first study was designed to test **H1** and **H2**. We proceeded to measure subjective responses to four

Male Participants				Female Participants			
Model	*M (SD)*	*Mdn*	*Mode*	Model	*M (SD)*	*Mdn*	*Mode*
Male Models							
Spartacos	3.42 (2.21)	2.5	2	Beach Boy	3.89 (1.76)	3	3
Beach Boy	4.03 (2.14)	3	3	Spartacos	4.33 (2.49)	5	1
Muchacho	4.58 (1.98)	5	5	Raphael	4.46 (2.52)	5	8
Raphael	4.62 (2.32)	4	4	Rob	4.59 (2.48)	5	2
Sol	4.69 (2.26)	5	7	Muchacho	4.63 (2.18)	4	4
Rob	4.77 (2.82)	5	8	Sol	4.63 (2.28)	5	2
Lee	4.85 (2.07)	5	4	Michael	4.72 (2.29)	5	4
Michael	5.04 (2.36)	6	6	Lee	4.74 (2.28)	5.5	7
Female Models							
Marie	3.69 (1.85)	4	4	Eastern Girl	3.91 (2.37)	4	4
Victoria 4.2	3.89 (1.99)	3	3	Girl Next Door 4	4.15 (2.27)	4	2
Girl Next Door 4	4.04 (2.29)	4	1	Rio	4.33 (2.31)	4	3
Stephanie 4	4.35 (2.97)	4.5	1	Amy	4.33 (2.27)	4	3
Asami	4.62 (2.48)	4	8	Marie	4.35 (2.41)	4	3
Amy	4.81 (2.08)	5.5	6	Asami	4.72 (2.25)	5.5	6
Eastern Girl	5.27 (2.15)	6	6	Victoria 4.2	4.83 (1.91)	5	5
Rio	5.35 (2.1)	5.5	7	Stephanie 4	5.37 (2.34)	6	8
All Participants							
Male Models				Female Models			
Model	*M (SD)*	*Mdn*	*Mode*	Model	*M (SD)*	*Mdn*	*Mode*
Girl Next Door 4	4.11 (2.26)	4	2	Beach Boy	3.94 (1.88)	3	3
Marie	4.14 (2.25)	4	3	Spartacos	4.04 (2.43)	4	1
Eastern Girl	4.35 (2.38)	4	4	Raphael	4.51 (2.44)	4.5	8
Amy	4.49 (2.21)	4	7	Muchacho	4.61 (2.11)	4.5	4
Victoria 4.2	4.53 (1.97)	5	5	Sol	4.65 (2.26)	5	2
Rio	4.66 (2.78)	4	3	Rob	4.65 (2.58)	5	8
Asami	4.69 (2.31)	4	6	Lee	4.78 (2.2)	5	7
Stephanie 4	5.04 (2.59)	6	8	Michael	4.83 (2.3)	5	7

Note. Models are listed in an ascending order by mean.

Table 1. Measures of Central Tendency Obtained for Pre-Test Models.

digital human models varying in human likeness. Human likeness was varied by manipulating the prototypicality of the models and the number of polygons constituting them. Emotional responses were measured by obtaining the participants' self-reports about the stimuli's eeriness, fear, disgust, and attractiveness.

3.1. Participants

164 undergraduate students (85 women, 79 men, M_{age} = 20.4, SD_{age} = 3.66) were recruited from Carleton University. Participants were compensated with bonus marks towards an introductory Psychology course.

3.2. Materials

3.2.1. Stimuli

A pretest was first conducted with the goal of selecting four digital models high in attractiveness to serve as the highest levels of human likeness. We assumed that attractiveness ratings would be sensitive to stimulus peculiarities, and therefore that selecting based on attractiveness would enhance our experimental control over feature congruency as well as the generalizability of our findings. 80 undergraduate participants (54 women, 26 men, M_{age} = 19.8, SD_{age} = 3.04) were recruited and compensated with bonus marks for participation. The stimuli consisted of 16 computer models (8 female and 8 male) purchased from DAZ 3D (n.d.) and Renderosity (n.d.) digital artist communities. These models included 3 "base" models, Victoria 4.2 (V4), Michael 4 (M4), and Stephanie 4 (S4), as well as a variety of derivatives that modified the facial morphology and textures of these base models. Blender software (2011) was used as a staging and rendering environment. Each model was staged using an identical three-point lighting arrangement and the image was rendered in a portrait style. Rendered images were cropped at the shoulders and then arranged into matrices separated by model sex.

Each face was presented with a label A-H and each matrix was accompanied by an 8-point attractiveness ranking scale. Participants were instructed to place the faces into a rank order from most attractive (1) to least attractive (8). Model rankings were then analyzed in terms of three measures of central tendency: mean, median, and mode. These results are presented in Table 1. For the male models, male and female participants' rankings identified Spartacos and Beach Boy as the most attractive. Hence, they were selected for the main experiment. The female model rankings were less consistent, but Girl Next Door 4 and Marie were ultimately selected. For both the male and female participants, Girl Next Door 4 had the lowest mode, and the lowest (or second lowest) median. The other selected model, Marie, obtained the lowest (or second lowest) mean and median ranking from the male participants. She also obtained the lowest (or second lowest) median and modal ranking. Although she ranked fifth in mean rankings, only .02 separated her from the third place. Therefore, Girl Next Door 4 and Marie appeared to be reasonable choices. Henceforth, the four selected models will identified as Male 1, Male 2, Female 1, and Female 2, respectively.

Fig 2. Stimuli Produced for Female 1, Experiment 1.

Using the selected models, 7 levels of prototypicality and 7 levels of geometric realism were then generated in the following manner. The highest level of each human likeness dimension was represented by the original models. Using Poser software (2010), the prototypicality of each model was reduced by simultaneously modifying eye size, mouth height, mouth size, face height, and eye separation. Specifically, eye size, mouth height, and mouth size were increased in increments of 12.5% to the maximum allowable value by the software; face height and eye separation were decreased in decrements of 12.5%. Using the resulting derivative models, geometric realism was manipulated using a polygon reducing algorithm in Blender. The number of polygons was reduced in decrements of 43.5%. As before, models were staged using an identical three-point lighting arrangement, rendered and then cropped at the

shoulders. Finally, in order to increase the symmetry of the resulting images, images were split vertically and the two halves were blended together at 50% opacity. In total, 49 images were created for each of the four models, for a total of 196 stimuli (see Figure 2). It was discovered after testing that Male 2 at prototypicality level 4 had been mistakenly generated using parameters for prototypicality level 3, and therefore level 4 is missing for this model from the procedure and the results that follow.

3.2.2. Scales and measures

Subjective reports were obtained for human likeness, eeriness, fear, disgust, and attractiveness using 7-point Likert scales. Statements were presented with the scales and worded according to the following format: "Please rate the extent to which you feel that the face shown above is [HUMAN-LIKE]". Each scale was anchored at the end-points and the mid-points by labels ("Not at all", "Moderately", and "Extremely") to facilitate consistent interpretation of the scale and scale intervals across participants.

3.3. Procedure

All testing, including stimulus presentation and the recording of participants' responses, was conducted using E-Prime 1.0 (Schneider, Eschman, & Zuccolotto, 2002a, 2002b). Individuals were told that they would be reporting their emotional reactions to digital human faces. The experimental design consisted of an incomplete block design. Stimuli were blocked according to the base model from which they were derived and according to model sex. Participants each received one block of male and one block of female stimuli. Counter-balancing and randomization procedures were used to control for any potential order effects. Blocks and the ordering of model sex were counter-balanced across participants. Within each block, the order of stimuli was randomized, and for each stimulus presentation the five scales were presented in a random order. Participants completed a total of 490 trials (2 blocks * 49 stimuli * 5 scales) in a single session lasting approximately 60 minutes.

3.4. Results

The data for one participant were lost due to a software malfunction and the data for another participant were removed because of a failure to comply with the task instructions. Hence, the data for 162 participants (78 men, 84 women, M_{age} = 20.4, SD_{age} = 3.67) were analyzed.

3.4.1. *The uncanny valley with subjective human likeness*

According to the uncanny valley theory, emotional responses to human-like stimuli are caused by stimulus placement along the continuum of human likeness. Thus, the theory predicts a non-linear relationship between human likeness and emotional response. We examined this possibility by following a common procedure observed in the uncanny valley literature to probe for the uncanny valley function (e.g., MacDorman & Ishiguro, 2006; MacDorman, 2006; MacDorman et al., 2009). Specifically, mean ratings for each of the stimuli were first collapsed across individuals. Next, eeriness and human likeness ratings were plotted separately for each of the four stimulus models. These plots are presented in Figure 3.

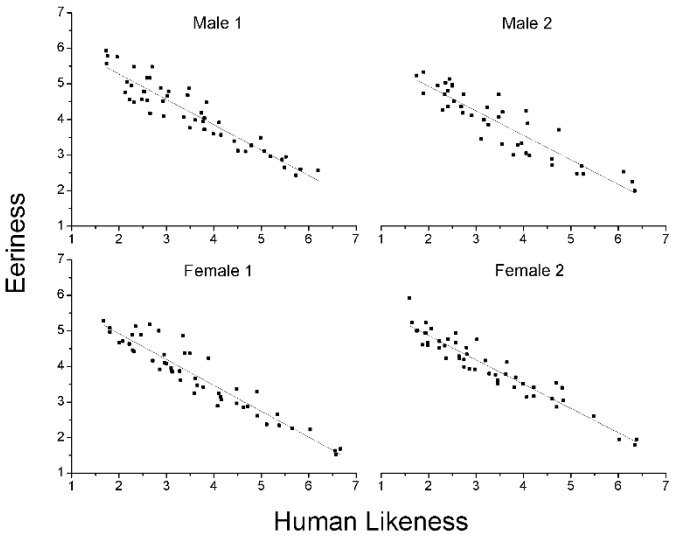

Fig 3. Mean Eeriness Ratings across Human Likeness, Aggregated by Stimulus and Separated by Stimulus Model.

It can be seen that the relationship between human likeness and eeriness appears to be linear for each of the four models. To test this interpretation of the data statistically, we fit linear, quadratic, and cubic functions to the data. If either a quadratic or cubic function were found to fit best, then either could reasonably be used as evidence to support the uncanny valley theory. The results of these curve estimations are presented in Table 2.

Set	Model	*RSS*	*AICc*	$\Delta_i(AIC)$	$w_i(AIC)$	R^2	*CI*
Female1*	Linear[1]	5.38	-103	0	.66	.89	.07
	Quadratic[2]	5.36	-101	2	.24	.89	-
	Cubic[3]	5.32	-99	4	.09	.89	-
Female2*	Linear	3.47	-124	0	.44	.91	.04
	Quadratic	3.43	-122	2	.19	.91	-
	Cubic	3.19	-124	0	.37	.91	-
Male1*	Linear	5.29	-104	0	.60	.88	.06
	Quadratic	5.20	-102	1	.30	.88	-
	Cubic	5.19	-100	4	.10	.88	-
Male2†	Linear	6.06	-73	0	.45	.83	.04
	Quadratic	5.75	-73	0	.42	.84	-
	Cubic	5.73	-71	2	.14	.84	-

[1]$K = 1$, [2]$K = 2$, [3]$K = 3$, *$n = 48$, †$n = 40$

Table 2. Model Comparison for Eeriness x Human Likeness, Separated by Stimulus Set.

The goodness-of-fit index we selected for model comparison was the Akaike Information Criterion (AIC; see Burnham & Anderson, 2002). We decided the AIC was most appropriate in these circumstances because we were comparing models of varying complexity and the AIC penalizes models with additional parameters (Burnham & Anderson, 2002). Thus, the AIC supports valid parsimony-based inferences. It has been suggested that the corrected formula should be used in the case of small sample sizes (i.e., when n/K < 40; see Burnham & Anderson, 2002). As the number of data points we were testing constitutes a small sample size, we used the corrected formula (*AIC_c = n*ln(RSS/n) + 2*K + (2*K*(K+1))/(n-K-1)*) in which RSS is the residual sum of squares. In addition, we calculated R^2 to examine the proportion of variance accounted for in the data. According to Table 2, the magnitudes of the linear and non-linear trends were large, accounting for a minimum of 83% of the variance in the data. It can also be seen, based on our goodness-of-fit measure, that the models best fit to the data were linear in each case. This is best illustrated by the Akaike Weights (w_i), which are a simple transformation of raw Akaike values that reflect the probability that a particular model among the set of models is correct (Wagenmakers & Farrell, 2004). Using these weights, evidence ratios can be obtained by dividing the weight of one model by the weight of another. All evidence ratios calculated were found in support of linear models. Moreover, with the exception of Male 2, linear functions were at least twice as likely as a quadratic function to be the best model for the data. It has also been recommended that candidate models be considered in context of a confidence set, which is akin to a confidence interval, and is defined as including all models greater than 10% of the highest Akaike Weight in the set (Royall, 1997). It might be noted that all models meet this minimum cut-off. Overall, these findings support our interpretation of the pattern of results in Figure 3 and are contrary to predictions made by the uncanny valley theory.

A limitation cited in previous research has been that not enough stimulus models were used (e.g., MacDorman et al., 2009). This is a concern because if stimuli are generated from a single base model, then observations might be driven by particular features which are unique to that model. The present experiment generated stimuli from four base models. These models were preselected based on attractiveness to control for anomalous features. Nevertheless, differences might still exist between the models. To preclude this possibility, pairwise comparisons were generated to examine

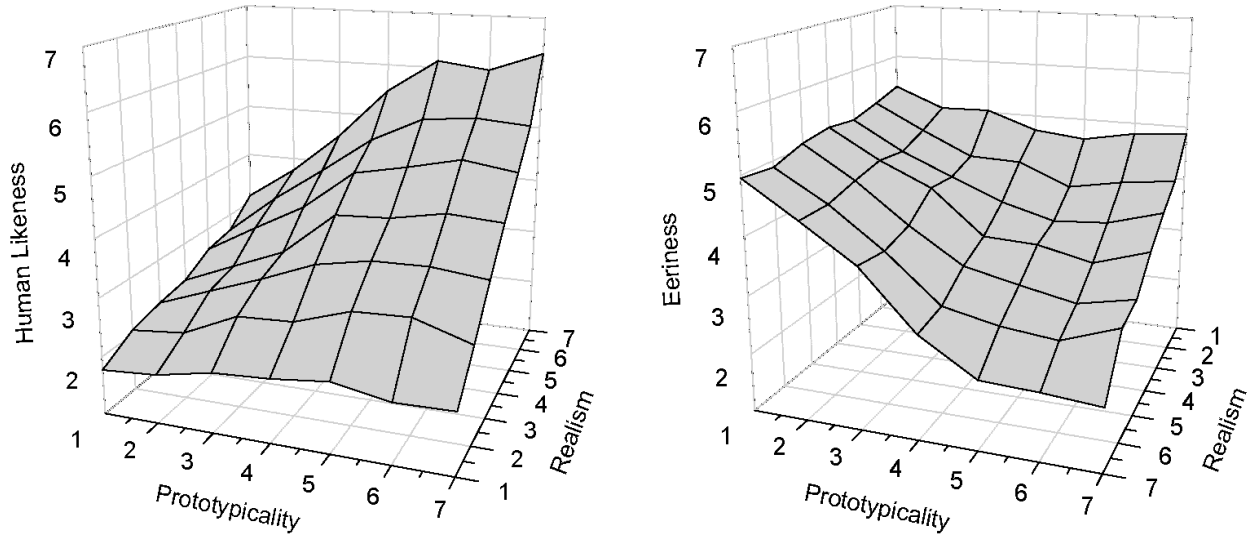

Fig 4. Surface Plots for Mean Human Likeness and Eeriness Ratings across Prototypicality and Realism Levels.

differences between mean ratings of the four stimulus models on each of the subjective rating scales. A t-test was calculated for each comparison and corrected using a Bonferroni adjustment for 30 tests (α_{adj} = .05 / 12 = .004). According to this criterion, none of the comparisons were significant. The smallest *p*-value obtained when comparing the attractiveness of Male 1 (*M* = 4.19, *SD* = 1.81) with Female 2 (*M* = 3.71, *SD* = 1.88) was *p* = .03, *t*(160) = 2.24. This comparison also produced the largest mean difference of .48 (Cohen's *d* = .26). Thus, these results suggest that although there might have been small differences among the models for some of the scales, there were no systematic differences that could lead to concerns about stimulus-specific confounds.

3.4.2. The uncanny valley with objective human likeness

Whereas the uncanny valley phenomenon is generally conceived in terms of subjective human likeness, some studies have observed the phenomenon when relating emotional response to objective levels of human likeness (e.g., Seyama & Nagayama, 2007). Although it has been established in the present experiment that a linear relationship exists between emotional response and subjective human likeness, it is possible that the relationship between emotional response and objective human likeness, as determined by levels of the stimulus parameters, might evidence an uncanny valley. Therefore, this possibility was examined by collapsing ratings across the stimulus models, and then plotting mean emotional responses across prototypicality and realism levels. These plots are presented in Figure 4.

According to the uncanny valley theory, when human likeness is related to eeriness, the resulting function should be cubic. In three dimensions, many different distributions could be found to support the uncanny valley, such as a parabola or a polynomial surface. Regardless of the particular shape that might be expected of an uncanny valley with three dimensions, the distribution of data would certainly be expected to deviate from a planar surface. However, the plots in Figure 4 suggest that this was not the case. Instead, it appears that bivariate linear relationships exist between the stimulus dimensions (i.e., prototypicality and realism) and subjective ratings. The shape of these distributions reflects a merger of two linear functions with the same direction and similar slopes. To assess this interpretation, a planar function was fit to these data. This function explained at least 79% of the variance. These findings suggest that bivariate linear functions provide more than adequate fits to the data sets. While it would undoubtedly be possible to fit more complex functions to the data, an examination of the surface plots presented in Figure 4 make it highly unlikely that any function fit would provide support for the uncanny valley.

3.4.3. Eeriness

Using self-reported fear, disgust, and attractiveness as proxies for the activation of fitness-relevant processes, we sought to clarify the phenomenology of eeriness. Specifically, we were interested in determining the relative importance of each of these emotions in accounting for eeriness. Ho et al. (2008) found that eeriness was related most closely to fear, while disgust and attractiveness were also found to play an important, though less significant, role. We expected to replicate the findings of Ho et al. (2008) in the present study.

In order to determine the relative importance of each emotion in relation to eeriness, we decided to calculate zero-order and squared semi-partial correlations. Zero-order correlations might provide a general indicator of the strength of these relationships, while a semi-partial correlation is particularly well suited to answer our question of relative importance as it represents the unique effect size for a specific predictor after taking into account other variables that are already predicting the dependent variable (Harlow, 2005; Howell, 2012). Therefore, we proceeded by running a simultaneous multiple regression with eeriness as the dependent variable, and fear, disgust, and attractiveness as independent variables. Together, these three variables accounted for 55% of the variance in eeriness. Using Cohen's (1988) rules of thumb, it might be noted that eeriness displayed a large zero-order correlation with fear (*r* = .71), disgust (*r* = .62), and attractiveness (*r* = -.52). The directions of these relationships are consistent with our predictions (H2). Fear produced a small squared semi-partial (r^2 = .11) when controlling for disgust and attractiveness, and thus it accounted for 11% of the unique variance in eeriness. Disgust produced a very small semi-partial (r^2 = .01) when controlling for fear and attractiveness, and thus it accounted for 1% of the unique variance in eeriness. Attractiveness also produced a very small semi-partial (r^2 = .02) when controlling for fear and disgust, and thus it accounted for 2% of the unique variance in eeriness. These results are consistent with Ho et al. (2008), and clearly indicate that eeriness is an emotion most closely related to fear. However, the unique variance accounted for by

attractiveness and disgust suggests that they also play a role, independently of fear and of one another.

3.5. Discussion

This experiment was designed to examine two hypotheses pertaining to the uncanny valley theory (MacDorman & Ishiguro, 2006; MacDorman et al., 2009; Mori, 1970). Although a number of studies have examined how stimulus placement on a continuum of human likeness impacts emotional response (Bartneck et al., 2007; MacDorman & Ishiguro, 2006; MacDorman, 2006; MacDorman et al., 2009; Schneider & Wang, 2007; Seyama & Nagayama, 2007; Steckenfinger & Ghazanfar, 2009), the findings of these studies have been inconsistent. We surmised that these inconsistencies were a result of differences in operational definitions of human likeness. Our greatest concern pertained to stimuli that confounded levels of human likeness by the inclusion of atypical or non-human features or category mergers (e.g., MacDorman & Ishiguro, 2006; Seyama & Nagayama, 2007). In the present experiment, these methodological issues were addressed by operationalizing human likeness as prototypicality and geometric realism. These choices ensured that the stimuli would vary sufficiently in their human likeness to find the uncanny valley, if it exists, without any other factors being present to influence the outcome. We predicted that (**H1**), a conservative operational definition of human likeness would produce a linear relationship with emotional responses. We observed that the relationship between emotional responses and human likeness were strongly linear. This was observed for all four of the emotions of interest: eeriness, fear, disgust, and attractiveness. Thus, we failed to find supporting evidence for Mori's (1970) uncanny valley theory. A potential concern that could be raised about these results is that a restricted range in the human likeness of the stimuli failed to capture the full cubic function predicted by the uncanny valley theory. An examination of Figure 1 indicates that two segments of the theoretical function are linear (between 0-60% and 80-100% human likeness). Therefore, if stimuli had elicited subjective human likeness ratings within either of these ranges, then the resulting function with any given emotion would have been linear. However, against this argument, it might be noted that participants made use of the full range of the human likeness and emotion scales.

As a secondary goal, we were interested in clarifying the phenomenology of eeriness. We expected to observe (**H2**) a positive relationship between eeriness and fear, and eeriness and disgust, and a negative relationship with attractiveness. We were interested in these relationships and what they might reveal about the uncanny valley phenomenon. Whereas an uncanny valley function was not obtained, the relationships we observed might still contribute to an account of eeriness as an emotional experience. Fear might be associated with threat avoidance or terror management, disgust with disease avoidance, and attractiveness with mate selection (Arndt, Greenberg, Pyszczynski, & Solomon, 1997; Thornhill & Gangestad, 1993). We observed that each of these three emotions were associated with eeriness, and appear to play a role, however eeriness appears to be primarily associated with fear. Therefore, it might be supposed that eeriness is most closely related to threat avoidance or terror management processes. Disease avoidance and mate selection might be involved, but their role appears to be modest. These findings are consistent with those obtained by Ho et al. (2008).

In sum, this experiment provided no support for the uncanny valley theory. Under conditions where it would predict non-linear patterns of emotional responses we instead observed linear patterns. It might be noted that individual-level data was not reported, but it was examined and found to be redundant with observations of linearity in aggregate data. Nevertheless, the fact remains that under certain conditions, human likeness continua do produce an uncanny valley pattern. Unfortunately, the extant literature has not been able to reliably produce this effect. Thus, in a second experiment, we sought to examine two conditions that we believed would generate an uncanny valley.

4. EXPERIMENT 2

As we argued previously, we believe that studies which have observed the uncanny valley phenomenon in the past might be accounted for in two ways: a category conflict or the inclusion of atypical features on otherwise typical human-like stimuli. In this experiment, we generated two continua of human likeness. One continuum consisted of the merger of two ontological categories, human and non-human animals. Based on the category conflict account, we predicted that (**H3**) a non-linear pattern would emerge across a continuum involving the merger of two ontological categories of objects. The second continuum varied human likeness within the ontological category of human by modifying texture realism. Across this continuum we introduced a highly atypical human feature. We hypothesized that (**H4a**) an unusual human feature placed on a stimulus high in human likeness would be perceived as more eerie than an unusual human feature placed on a stimulus low

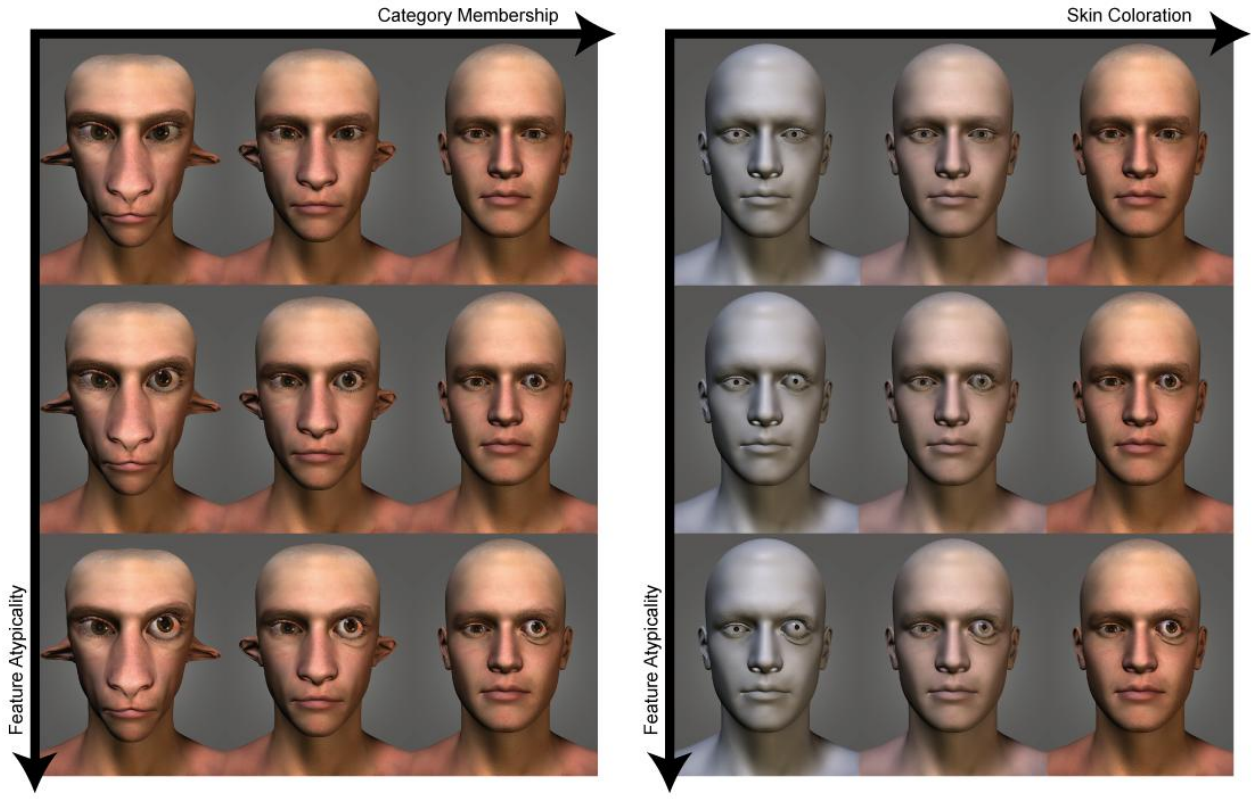

Fig 5. Sample stimuli produced for Experiment 2.

in human likeness. We also hypothesized that (**H4b**) feature atypicality would additively interact with human likeness in predicting eeriness, such that the combined effect of human likeness and feature atypicality would elicit more eeriness than either effect alone. The hypothesized direction of this additive effect will be specified depending on **H4a**.

4.1. Participants

47 undergraduate students (28 women, 19 men, M_{age} = 21.43, SD_{age} = 5.93) were recruited for this experiment. They were compensated with bonus marks towards an introductory Psychology course.

4.2. Materials

4.2.1. Stimuli

Two of the four models that served in Experiment 1 (M4 and V4) were used again in Experiment 2. For each, two continua of human likeness were produced: a continuum of prototypicality and a continuum of texture realism. Examples of the stimuli are shown in Figure 5. Prototypicality was generated as a sequence of anatomical changes between human and non-human animal prototypes. Specifically, we used the Creature Creator add-on for Poser (2010) to produce our animal prototype, by modifying several head-shape morph parameters in order to create a goat-like appearance[1]. Thus, the animal prototype corresponded to maximum values on these parameters, and the human prototype corresponded to no value for these parameters. Each level on the continua was then generated in 7 steps between the maximum and minimum values (i.e., increases of 16.7% in parameter values), creating a total of 7 levels of prototypicality. The texture realism continua were produced by creating an end-point for the model in which the skin texture was replaced by a neutral grey colour. The grey texture and the original texture were then blended together in 7 steps (e.g., 0/100, 16.7/83.3, 33.3/66.7, 50/50, etc…) to create a total of 7 levels of realism. We generated an atypical feature by modifying the texture, size, and orientation of the models' left eye such that it appeared to have been "rolled back". These changes were also produced incrementally for 7 levels, and these levels were crossed with each of the human likeness continua. In all, 196 stimuli were produced (2 models * 2 continua * 7 levels of human likeness * 7 levels of feature atypicality).

[1] Specifically, we created a Master parameter consisting of the Goat (0.4), AlienGrey (0.2), and NoseCat (1.0) morphs from the Creature Creator add-on, and the MouthSmileOpen (-0.06) from the standard base morphs. Values indicated in parentheses correspond to the maximum value set for the Master parameter.

4.2.2. Scales and measures

Subjective reports were obtained for human likeness, eeriness, and pleasantness using 7-point Likert scales. Pleasantness was measured in addition to eeriness as positive and negative responses have resulted in slightly different functions in the past (e.g., see "eeriness" vs. "familiarity" ratings in MacDorman & Ishiguro, 2006), and we were therefore interested in the possibility that positive affect might provide unique information bearing on our hypotheses. Scales were presented in the same manner as in Experiment 1.

4.3. Procedure

All testing, including stimulus presentation and the recording of participants' responses, was conducted using E-Prime 1.0 (Schneider, Eschman, & Zuccolotto, 2002a, 2002b). Individuals were told that they would be reporting their emotional reactions to digital human faces. Stimuli were blocked according to model sex and type of human likeness continua. Thus, there were four blocks of 49 stimuli each. Counter-balancing and randomization procedures were used to control for any potential order effects. The order of blocks was counter-balanced across participants. Within each block, the order of stimuli was randomized, and for each stimulus presentation the three scales were presented in a random order. Participants completed a total of 588 trials (4 blocks * 49 stimuli * 3 scales) in a single session lasting approximately 60 minutes.

4.4. Results

We proceeded to examine hypotheses (**H3** and **H4a**/**H4b**) as in Experiment 1, by first examining emotional response in relation to subjective human likeness, and then, in relation to objective human likeness (as defined by the stimulus dimensions). We collapsed mean ratings for each of the continua across individuals, and then examined the shapes of the distributions.

4.4.1. The uncanny valley with subjective human likeness

To examine the relationship between subjective human likeness and eeriness, we generated plots of these ratings separately for the prototypicality and realism continua. These plots are presented in Figure 6.

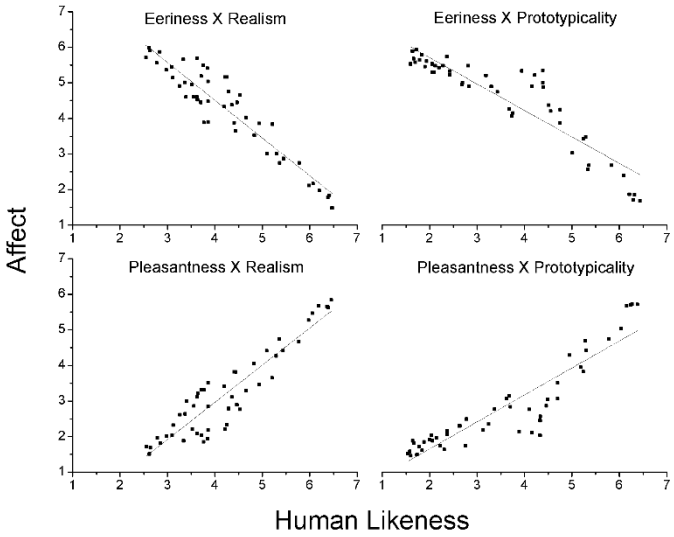

Fig 6. Mean Emotion Ratings across Human Likeness.

Figure 6 suggests that the relationship between human likeness and eeriness was linear in the realism condition. Although the trends in the prototypicality condition appear to be generally linear, it should be noted that there is a cluster of data points that deviate the human likeness axis. We next proceeded to estimate the curves of these functions, the results of which are presented in Table 3.

As in Experiment 1, our goodness-of-fit measure was the corrected version of the Akaike Information Criterion. It might be noted that the Akaike Weights provide an indication of the relative probability of a model's accuracy, and that they might be considered in context of a confidence set (defined as including all models greater than 10% of the highest Akaike Weight in the set; see Royall, 1997). Table 3 reports the interval values that might be used for examining each set. Based on this criterion, it is interesting to note that in both realism conditions, all models are candidates when examining both eeriness and pleasantness ratings. Yet, in the prototypicality conditions, the linear models are not candidates as their Akaike Weights do not fall within the specified confidence intervals. The R^2 values obtained are consistent with this observation. Specifically, in the Eeriness x Realism and Pleasantness x

Set	Model	*RSS*	*AICc*	$\Delta_i(AIC)$	$w_i(AIC)$	R^2	*CI*
Eerieness x Realism*	Linear[1]	8.76	-82	4	.11	.88	.06
	Quadratic[2]	7.79	-86	0	.65	.89	-
	Cubic[3]	7.75	-84	2	.24	.89	-
Eerieness x Prototypicality*	Linear	12.69	-64	28	.00	.83	.07
	Quadratic	7.10	-90	1	.34	.90	-
	Cubic	6.60	-92	0	.66	.91	-
Pleasantness x Realism*	Linear	10.45	-74	4	.10	.85	.06
	Quadratic	9.25	-77	0	.65	.87	-
	Cubic	9.18	-76	1	.25	.87	-
Pleasantness x Prototypicality*	Linear	12.26	-66	37	.00	.84	.09
	Quadratic	5.92	-99	4	.13	.92	-
	Cubic	5.23	-103	0	.87	.93	-

[1]$K = 1$, [2]$K = 2$, [3]$K = 3$, *$n = 49$

Table 3. Model Comparison for Eeriness x Human Likeness and Pleasantness x Human Likeness.

Realism conditions, the unique variance accounted for the specified confidence intervals. The R^2 values obtained are consistent with this observation. Specifically, in the Eeriness x Realism and Pleasantness x Realism conditions, the unique variance accounted for by a quadratic or cubic model beyond a linear model appears to be negligible. In contrast, the quadratic and cubic models explain 7 to 9% more variance than a linear model in the Eeriness x Prototypicality and Pleasantness x Prototypicality conditions.

The two conditions involving the prototypicality factor are statistically non-linear, and our visual impression of these trends suggests that this outcome might have been caused by a cluster of outlying data points. Therefore, we decided to examine these data points more closely. To this end, a separate linear regression was conducted in the prototypicality condition for both eeriness and pleasantness, with human likeness as the predictor in each case.

The two conditions involving the prototypicality factor are statistically non-linear, and our visual impression of these trends suggests that this outcome might have been caused by a cluster of outlying data points. Therefore, we decided to examine these data points more closely. To this end, a separate linear regression was conducted in the prototypicality condition for both eeriness and pleasantness, with human likeness as the predictor in each case. The distance between each data point and the regression model was output as standardized residual values. We examined data points with extreme residual values. In the eeriness regression, a total of 7 data points were identified with residual values greater than 1.5 *SD* from the regression model, and 6 out of these 7 were found between 3.75 and 4.5 on the human likeness axis. Moreover, 4 data points exceeded a distance of 2 *SD*, and all of these data points fell within this same range. It should be noted that these data points correspond to the outlying data points in Figure 6.

In the pleasantness regression, a total of 8 data points were identified with residual values greater than 1.5 *SD* from the regression model, and 7 out of 8 were found between 3.75 and 4.5 on the human likeness axis. Of these 7 data points, 2 exceeded a distance of 2 *SD*. Thus, it appears that around the mid-point of the human likeness axis, eeriness and pleasantness ratings were an exception to the overall trend. Further, it is likely that these points were responsible for the statistical non-linearity outcomes reported above. As this finding is consistent with our *a priori* prediction, it supports our category merger hypothesis (**H3**).

4.4.2. The uncanny valley with objective human likeness

Next, we proceeded to examine the relationship between objective human likeness and emotional response. First, we expected to observe a non-linear relationship between eeriness and stimulus dimensions along the prototypicality continuum (**H3**). Secondly, we also expected that an unusual human feature would elicit greater eeriness at a high level of human likeness than at a lower level (**H4a**). Finally, we anticipated that the salience of an unusual feature would additively interact with human likeness in predicting eeriness (**H4b**). Before examining these hypotheses, it was necessary to first check that the stimuli elicited the expected pattern of human likeness ratings across stimulus levels. As each of the conditions consisted of two stimulus parameters, we examined ratings in three-dimensional space. Human likeness plots are presented in the top panels of Figure 7.

We expected to find planar surfaces, and Figure 7 suggests that this expectation was satisfied. That is, levels of the stimulus dimensions corresponded to approximately monotonic changes in human likeness. To confirm this interpretation, a planar function was fit to the data in each condition, and was found to explain at least 91% of the variance.

We then produced a plot of eeriness ratings across the prototypicality continuum in three-dimensional space to examine **H3** and **H4a**. We expected to find areas along the surface that would not correspond to the overall planar trend (e.g., peaks and troughs in the surface). For example, **H3** would predict that eeriness ratings would deviate from the planar trend mid-way along the prototypicality continuum, whereas **H4a** might predict that eeriness ratings would deviate from the planar trend at high levels on both human likeness and feature atypicality. However, against our expectations, the

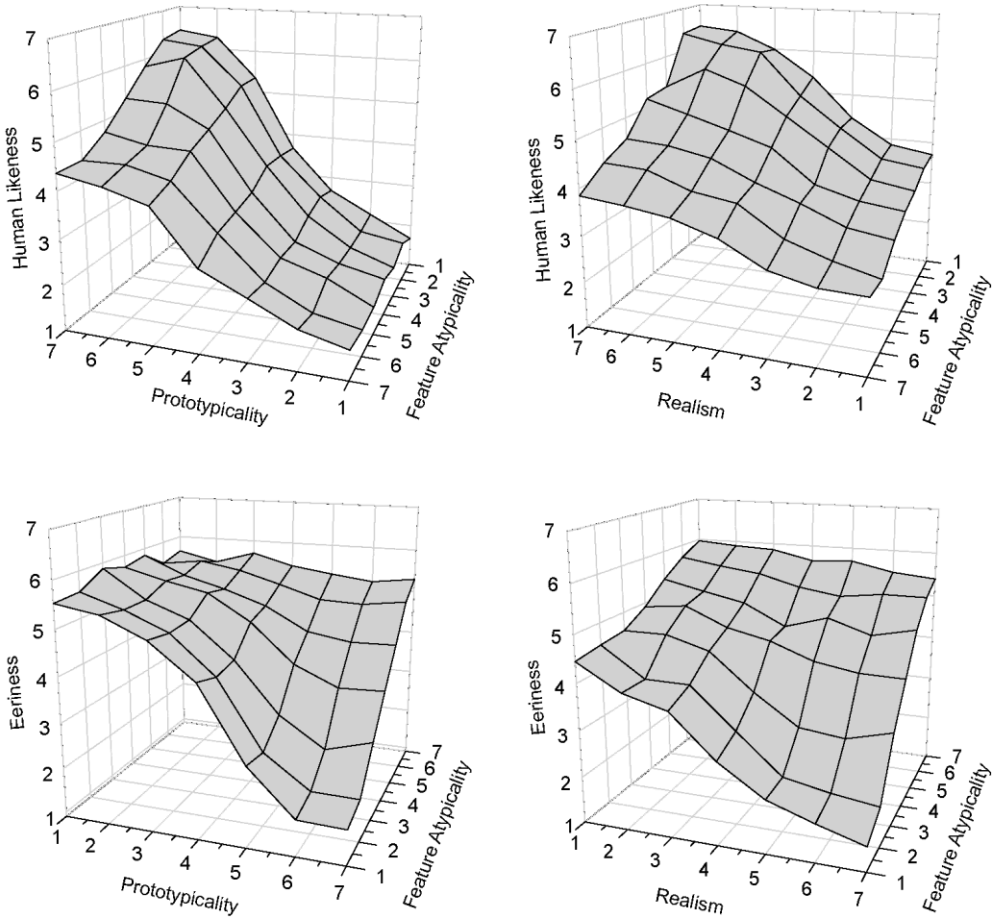

Fig 7. Surface Plots for Mean Human Likeness and Eeriness Ratings across Prototypicality x Feature Atypicality, and Realism x Feature Atypicality.

bottom panels of Figure 7 suggests that these surfaces were planar and that levels of the stimulus dimensions elicited approximately monotonic changes in eeriness. This was confirmed by fitting a planar function to the data, which explained a minimum of 77% of the variance in the data.

Finally, we examined **H4b** by testing the interaction between feature atypicality and human likeness on eeriness ratings using a hierarchical regression approach. Separate regressions were performed for each of the human likeness continua. First, we examined the prototypicality condition. Eeriness was entered as a dependent variable, and feature atypicality and prototypicality were entered as independent variables in the first step, followed by their product term in the second step. The first step was found to be significant ($F(2,57) = 76.05$, $p < .05$, $R^2 = .77$). In addition, the second step was also significant, ($F(3,70) = 247.61$, $p < .05$, $R^2 = .94$), and the interaction accounted for an additional 17% of the variance in ratings. We then examined the realism condition in the same manner, entering eeriness as the dependent variable, feature atypicality and realism as predictors in the first step, and their product term in the second step. Again, we observed that the first ($F(2,64) = 195.93$, $p < .05$, $R^2 = .90$) and second ($F(3,70) = 414.74$, $p < .05$, $R^2 = .96$) steps were significant. The interaction term accounted for only 6% additional variance in this condition, however.

Nevertheless, the specific nature of the interactions must be taken into consideration. Our hypothesis **H4b** predicts an additive interaction for which the joint influence of feature atypicality and human likeness on eeriness should be greater than the influence of either feature considered individually. More specifically, we expected that the effect of feature atypicality would increase as human likeness decreased because support for **H4a** was not found. The eeriness plots in Figure 7 supports this conjecture. In the realism condition it might be observed, for example, that the combined effect of minimum realism and maximum feature atypicality was more eerie ($M = 5.97$, $SD = 2.07$) than either minimum realism ($M = 5.40$, $SD = 1.58$, $t(186) = 2.12$, $p = .02$) or maximum feature atypicality ($M = 4.45$, $SD = 1.68$, $t(186) = 5.53$, $p < .001$) alone. An additive interaction was also observed in the prototypicality condition. At the mid-point along each stimulus dimension, for example, the

combined effect of prototypicality and feature atypicality was more eerie (M = 4.70, SD = 1.51) than either prototypicality (M = 4.22, SD = 1.87, $t(186) = 1.94$, $p = .03$) or feature atypicality (M = 3.44, SD = 1.86, $t(186) = 5.10$, $p < .001$) alone. The slope of the plane observed in the realism condition suggests that the joint influence of realism and feature atypicality was consistent across these dimensions, whereas the surface in the prototypicality condition suggests that only particular levels (such as the mid-point) produced an additive interaction. These results provide support for **H4b**.

4.5. Discussion

Experiment 2 was designed to test two predictions regarding the uncanny valley phenomenon. Specifically, we hypothesized (**H3**) that if the phenomenon could be attributed to a category conflict aroused by categorical ambiguity of a stimulus, and if human likeness is operationalized as a merger of two categories, then a non-linear function would emerge across the continuum. Second, we hypothesized (H4) that if the uncanny valley phenomenon could be attributed to the inclusion of an atypical feature across a continuum of human likeness, then eeriness would be maximized when human likeness was greatest and an atypical feature was most salient (H4a). We also expected that feature atypicality and human likeness would display an additive interaction with respect to eeriness ratings (**H4b**). In order to test these hypotheses, we generated two continua of human likeness, one in which stimuli ranged between human and animal prototypes, and a second in which human likeness was operationalized as texture realism. Each continuum also consisted of a second stimulus dimension pertaining to feature atypicality.

Our results provided support for our category merger hypothesis (**H3**). A close examination of eeriness ratings within the prototypicality condition suggested a cluster of data points that were not predicted by a linear function. Consistent with our *a priori* prediction, these data points showed heightened levels of eeriness at approximately the mid-point of subjective human likeness. We did not find support for **H4a**, however. We had predicted that an unusual feature would elicit more eeriness at higher levels of stimulus human likeness than at lower levels. The direction of the joint influence of feature atypicality and human likeness was such that greater feature atypicality, combined with less human likeness, resulted in higher eeriness ratings (**H4b**). This latter finding might be consistent with MacDorman et al.'s (2009) findings. They observed that the joint effect of an atypical feature and low texture realism on eeriness was greater than either effect alone. They also found, however, that the effect of an atypical feature was greater at higher levels of texture realism than at lower levels, which is inconsistent with our present findings. It might be noted that there were differences in the atypical feature that was used, as MacDorman et al. (2009) increased the eye size of models by 50%, whereas we modified the texture, size, and orientation of the models' left eye such that it appeared to have been "rolled back". However, at this time we are unable to offer any satisfying explanation as to why these features might have led to different outcomes. Nevertheless, we would like to speculate that certain kinds of atypical features might elicit a kind of category conflict if they appear to be incoherent with the object to which they belong. For example, if an atypical feature appears on, but cannot realistically be expected to belong to, a human, then the stimulus as a whole is incoherent. This overall stimulus incoherence might elicit cognitive dissonance if the atypical feature belongs to another category of objects (e.g., a robot), or if the stimulus merely appears to be non-human. This idea is reinforced by previous studies which found an uncanny valley. Consider again the stimuli produced in MacDorman and Ishiguro (2006). Their morphing procedure created stimuli for which certain features were transposed from one image to the other. For instance, in the introduction, we described how the black dot present on the robot image was superimposed on the human-like images as the sequence progressed towards the human end of the continuum. Ostensibly, this feature did not belong on a human, yet its identity or categorical provenance was unclear. Finally, it should be added that MacDorman and Ishiguro's stimuli had other transposed features such as the robot's chassis which appeared to be attached to the human-like figure's back. Our original conclusion was that the atypicality of one or more of these features had resulted in different possible stimulus interpretations. Yet, it remains to be seen whether the uncanny valley that was observed was due to an ambiguous non-human feature such as the black dot, or a non-human feature that can be identified as belonging to another category such as the robot chassis. It remains possible that, under either condition, an uncanny valley might emerge.

5. General Discussion

In two studies, we examined the relationship between human likeness and eeriness using digital human faces. In the first study, we generated continua of human likeness that varied in terms of stimulus realism and facial proportions, while controlling for extraneous

variation. Across both continua linear patterns of emotional responding were observed. Thus, contrary to the classic interpretation of the uncanny valley theory (Mori, 1970), stimuli defined by a near-perfect resemblance to humans do not appear to cause people to experience greater negative affect relative to when they have perfect human likeness or little to no human likeness.

Based on our reading of the literature, we expected that an uncanny valley would emerge under one of two conditions. First, we speculated that introducing an atypical feature across a continuum of human likeness might cause stimuli higher in human likeness to be judged more critically than stimuli lower in human likeness, as might be expected given evolved mechanisms for threat detection in conspecifics. This *atypical feature* hypothesis was not supported. Second, we speculated that if human-like stimuli varied in terms of category membership, then stimuli located at a mid-point between two categories could elicit negative affect due to conflicting stimulus interpretations. Support for this *category conflict* hypothesis was obtained. Taken together, these findings suggest that previous evidence obtained in support of the uncanny valley theory (MacDorman & Ishiguro, 2006; Mitchell et al., 2011; Saygin et al., 2011; Seyama & Nagayama, 2007) might be accounted for on the basis of the stimulus belonging simultaneously to multiple ontological categories, which elicits a state of discomfort because it is ambiguous and conflicting. Although we can only speculate as to the specific mechanisms underpinning this effect, we suggest that future research might find answers in neuroimaging studies. For example, conflict resolution processes are thought to reside in the anterior cingulate (Botvinick et al., 2001), and at least one neuroimaging study found that a state of cognitive dissonance engaged dorsal anterior cingulate and anterior insula regions (van Veen, Krug, Schooler, & Carter, 2009). Thus, it might be expected that stimuli associated with category conflict would engage these same regions.

An important point concerns the external validity of our findings. Our purpose was descriptive insofar as we were interested in establishing empirically the existence of a phenomenon. However, it remains to be seen whether our findings hold across other contexts in which human-like stimuli are presented, with other stimuli that represent human-nonhuman mergers, and with other populations of individuals. It might be expected that affective responses to human-like stimuli would vary according to past experience with similar stimuli, and that experience might be related to cultural or generational differences. Contextual and population differences are certainly an interesting avenue for future research. If eeriness is a consequence of category conflict, then presumably the degree to which a conflict occurs would be moderated by individual differences in category representations, particularly with respect to category boundaries (i.e., conflict would be strongest when categories are represented as mutually-exclusive).

Based on our findings, we would like to make the following recommendations to designers. We believe that a fear of the uncanny valley is unwarranted, and even potentially detrimental to the pursuit of design goals where human likeness is involved. This is especially true for advances in human-likeness that do not introduce non-human features, as is the case with the advance of resolution in motion capture technology, the resolution of graphical textures, or the polygon count of computer-generated models. Under these conditions, greater realism is likely to improve the experience. However, given that human-nonhuman category mergers can elicit negative responses, caution might still be advised where the possibility exists for features diagnostic of a non-human category to appear. For example, in the design of human-like robots, it would be unwise to present a near-perfect human-like visual appearance with distinctly robotic voice. Insofar as modifications to human likeness do not introduce changes in category membership, then we suggest that the risk of an uncanny valley phenomenon occurring is minimal. Designers face the great challenge of re-creating that which is most familiar to us – the human likeness – this task will not be easy because expectations are so high, but it is an undertaking with the potential for equally great social and economic rewards.

References

Arndt, J., Greenberg, J., Pyszczynski, T., & Solomon, S. (1997). Subliminal exposure to death-related stimuli increases defense of the cultural worldview. *Psychological Science*, 8(5), 379–385. doi:10.1111/j.1467-9280.1997.tb00429.x

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is the uncanny valley an uncanny cliff? *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 368–373). doi:10.1109/ROMAN.2007.4415111

Blender. (2011). Blender Foundation.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652. doi:10.1037/0033-295X.108.3.624

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Buss, D. M. (2005). *The handbook of evolutionary psychology*. John Wiley & Sons.

Cameron, J. (2009). *Avatar.*

Chaminade, T. D., Hodgin, J. K., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, 2(3), 206–216.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (2nd ed., pp. 91–115). New York: Guilford.

DAZ 3D. (n.d.). DAZ Productions, Inc. Retrieved from http://daz3d.com/

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, 67(3), 382–394. doi:10.1037/0022-3514.67.3.382

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, USA: Stanford University Press.

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications*, 28(4), 11–17. doi:10.1109/MCG.2008.79

Hanel, M. (2008, June 11). Transsiberian: Trains, pain and creepiness. *Vanity Fair*. Retrieved from http://www.vanityfair.com/online/daily/2008/07/transsiberian-trains-pain-and-creepiness

Harlow, L. L. (2005*). The essence of multivariate thinking: Basic themes and methods*. Psychology Press.

Ho, C.-C., Pramono, Z. A. D., & Macdorman, K. F. (2008). Human emotion and the uncanny valley: A GLM, MDS, and isomap analysis of robot video ratings. *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*. doi:10.1145/1349822.1349845

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Cengage Learning.

Keegan, R. W. (2009). *The futurist: The life and films of James Cameron*. Random House of Canada.

Kloc, J. (2009, June 7). Into the uncanny valley. *Seed Magazine*. Retrieved from http://seedmagazine.com/content/article/uncanny_valley

Levy, S. (2004). Why Tom Hanks is less than human: While sensors can't capture how humans act, humans can give life to digital characters. *Newsweek*, 144(21).

Loder, K. (2004, November 10). "The Polar Express" is all too human. *MTV News*. Retrieved from http://www.mtv.com/news/articles/1493616/kurt-loder-on-polar-express.jhtml

MacDorman, K F, & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3), 297–337. doi:10.1075/is.7.3.03mac

MacDorman, Karl F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *Proceedings of the ICCS/CogSci-2006: Toward Social Mechanisms of Android Science*. Vancouver.

MacDorman, Karl 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:10.1016/j.chb.2008.12.026

Mitchell, W. J., Sr, K. S., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1), 10–12. doi:10.1068/i0415

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

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior*, 26(1), 88–105. doi:10.1016/j.evolhumbehav.2004.08.002

Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with disabilities. *Journal of Nonverbal Behavior*, 27(2), 65–87. http://dx.doi.org/10.1023/A:1023910408854

Pollick, F. E. (2009). The search for the uncanny valley. In P. Daras & O. M. Ibara (Eds.), *UC Media 2009* (Vol. 40, pp. 69–78). Venice, Italy: Springer. doi:10.1007/978-3-642-12630-7

Poser. (2010). Smith Micro Software.

Renderosity. (n.d.). Renderosity. Retrieved from http://www.renderosity.com

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 57, 199–226. doi:10.1146/annurev.psych.57.102904.190208

Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm*. Chapman & Hall.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7(4), 413. doi:10.1093/scan/nsr025

Schneider, E., & Wang, Y. (2007). Exploring the uncanny valley with japanese video game characters. *Proceedings of DiGRA 2007: Situated Play*, 546–549.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools Inc.

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, 16(4), 337–351. doi:10.1162/pres.16.4.337

Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences*, 106(43), 18362. doi:10.1073/pnas.0910063106

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature*, 4(3), 237–269. doi:10.1007/BF02692201

Tinwell, A., & Grimshaw, M. (2009). Bridging the uncanny: an impossible traverse? *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, 66–73. doi:10.1145/1621841.1621855

van Veen, V., Krug, M. K., Schooler, J. W., & Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nature Neuroscience*, 12(11), 1469–74. doi:10.1038/nn.2413

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196. doi:10.3758/BF03206482

Walters, G. D. (2004). Understanding the popular appeal of horror cinema: An integrated-interactive model. *Journal of Media* Psychology, 9(2).

Zemeckis, R. (2004). *The Polar Express*. Warner Bros. Pictures.