Cognition 146 (2016) 190–205

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not

Karl F. MacDorman *, Debaleena Chattopadhyay

Indiana University School of Informatics and Computing, 535 West Michigan St., Indianapolis, IN 46202, USA

ARTICLE INFO

Article history:
Received 23 December 2014
Revised 21 September 2015
Accepted 22 September 2015
Available online 1 October 2015

Keywords:
Anthropomorphism
Computer animation
Face perception
Photorealism

ABSTRACT

Human replicas may elicit unintended cold, eerie feelings in viewers, an effect known as the uncanny valley. Masahiro Mori, who proposed the effect in 1970, attributed it to inconsistencies in the replica's realism with some of its features perceived as human and others as nonhuman. This study aims to determine whether reducing realism consistency in visual features increases the uncanny valley effect. In three rounds of experiments, 548 participants categorized and rated humans, animals, and objects that varied from *computer animated* to *real*. Two sets of features were manipulated to reduce realism consistency. (For humans, the sets were eyes–eyelashes–mouth and skin–nose–eyebrows.) Reducing realism consistency caused humans and animals, but not objects, to appear eerier and colder. However, the predictions of a competing theory, proposed by Ernst Jentsch in 1906, were not supported: The most ambiguous representations—those eliciting the greatest category uncertainty—were neither the eeriest nor the coldest.

1. Introduction

1.1. The problem of the uncanny valley and its significance

The uncanny valley hypothesis predicts a negative emotional appraisal of human replicas that appear or behave not quite human. This appraisal is typically attributed to the perceived salience of their nonhuman features (MacDorman, Srinivas, & Patel, 2013; Mitchell et al., 2011). Mori (1970/2012) graphs the uncanny valley as a dip in an otherwise positive relation between human likeness and affinity. He illustrates the concept with androids, moving corpses and mannequins, prosthetic arms, and a robot whose smile turns creepy when formed at half speed. All these examples combine human and nonhuman features, causing cold, eerie feelings in viewers. More recently scholars and film critics have related the uncanny valley to viewers' dyspathy for three-dimensional (3D) computer-animated heroes that closely resemble real people (Butler & Joschko, 2009; Freedman, 2012; cf. *Mars Needs Moms*, 2011, *A Christmas Carol*, 2009, *Beowulf*, 2007, *The Polar Express*, 2004, and *Final Fantasy: The Spirits Within*, 2001). Some viewers fail to identify with the characters, experiencing them instead as soulless or vacant.

Although the uncanny valley has been empirically investigated since 2005, the hypothesis remains controversial. Some studies have been interpreted as supporting the hypothesis (e.g., in human adults, Burleigh, Schoenherr, & Lacroix, 2013; Ho, MacDorman, & Pramono, 2008; MacDorman, Green, Ho, & Koch, 2009; Mitchell et al., 2011; Seyama & Nagayama, 2007; Tinwell, Grimshaw, & Abdel Nabi, 2015; Tinwell, Grimshaw, Abdel Nabi, & Williams, 2011; in children, Tinwell & Sloan, 2014; in infants, Lewkowicz & Ghazanfar, 2012; Matsuda, Okamoto, Ida, Okanoya, & Myowa-Yamakoshi, 2012; and in other primates, Steckenfinger & Ghazanfar, 2009). Other studies have instead been interpreted as supporting alternative hypotheses (e.g., an uncanny cliff, Bartneck, Kanda, Ishiguro, & Hagita, 2007, or wall, Tinwell, Grimshaw, & Williams, 2011; uncanniness caused by attributions of experience, Gray & Wegner, 2012; or a correlation between perceptual discrimination difficulty and positive affect, Cheetham, Suter, & Jäncke, 2014). Still other studies have found support for the uncanny valley hypothesis but not for Mori's (1970/2012) hypothesis that movement amplifies its effect (Piwek, McKay, & Pollick, 2014; Thompson, Trafton, & McKnight, 2011).

Theories to explain the uncanny valley are not lacking. They range from the biological to the cultural (reviewed in MacDorman, Green, et al., 2009; MacDorman & Ishiguro, 2006a, 2006b; MacDorman, Vasudevan, & Ho, 2009; Pollick, 2010). Rather, evidence is insufficient to decide among them. For example, Mori (1970/2012) proposed that the uncanny valley effect is a survival instinct, an aversive response to proximal threats like dead or diseased bodies and dangerous species of animals. We have further developed Mori's survival instinct hypothesis by relating the

* Corresponding author.
E-mail address: kmacdorm@indiana.edu (K.F. MacDorman).

http://dx.doi.org/10.1016/j.cognition.2015.09.019
0010-0277/

uncanny valley to the detection and avoidance of potential vectors of infection or of infertile or less fit mates (e.g., Neanderthals, MacDorman, Green, et al., 2009; MacDorman & Ishiguro, 2006a, 2006b; MacDorman, Vasudevan, et al., 2009; Moosa & Ud-Dean, 2010). We have also proposed the uncanny valley may result from inconsistency in the realism of an anthropomorphic entity's features (MacDorman, Green, et al., 2009; MacDorman, Vasudevan, et al., 2009). Perhaps the earliest explanation of uncanniness in human-looking entities is category uncertainty, in particular uncertainty about whether an entity is real or human (Jentsch, 1906/1997).

This study experimentally examines whether the uncanny valley effect is increased by category uncertainty (Burleigh & Schoenherr, 2014; Burleigh et al., 2013; Green, MacDorman, Ho, & Vasudevan, 2008; Jentsch, 1906/1997; Kang, 2009; MacDorman, Vasudevan, et al., 2009; Yamada, Kawabe, & Ihaya, 2013) or, as an alternative theory, by realism inconsistency (MacDorman, Green, et al., 2009; MacDorman, Vasudevan, et al., 2009; Meah & Moore, 2014; Mitchell et al., 2011; Moore, 2012). Category uncertainty denotes an inability to determine the category to which an entity belongs, such as whether a face is that of a real human being or a 3D computer model. Realism inconsistency denotes a mismatch in the realism of an entity's features, such as some facial features appearing human and others nonhuman (e.g., computer-animated skin paired with real eyes and mouth). We chose to evaluate category uncertainty and realism inconsistency theories because of their prominence in the literature on the uncanny valley.

1.2. Category uncertainty theories

1.2.1. Categorical perception

Although Mori proposed the uncanny valley in 1970, Jentsch, as early as 1906, developed a theory identifying category uncertainty as the cause of uncanniness (Jentsch, 1906/1997; MacDorman & Ishiguro, 2006a). He asserts that eerie feelings are most reliably elicited by uncertainty about whether an entity is inanimate or animate, or whether it is nonhuman or human. Category uncertainty occurs whenever an entity transitions from one category to another, qualitatively distinct category by a quantitative metric—for example, a *fertilized ovum* transitioning to *a person* by the metric *developmental chronology* (Ramey, 2005). Mori's (1970/2012) graph depicts *industrial robot* transitioning to *healthy person* by the metric *human likeness*. Owing to categorical perception, small changes along the continuum between two categories should appear much larger than equal-sized changes within either category (Beale & Keil, 1995; Campbell, Pascalis, Coleman, Wallace, & Benson, 1997; Etcoff & Magee, 1992; Harnad, 1987; Iverson & Kuhl, 1995). This phenomenon is also known as the perceptual magnet effect (Feldman, Griffiths, & Morgan, 2009). Near the category boundary, the increased salience of these changes could make them jarring. Categorical perception has been found on a continuum from 3D computer models to photographs of real people (Cheetham, Pavlovic, Jordan, Suter, & Jäncke, 2013; Cheetham, Suter, & Jäncke, 2011; Looser & Wheatley, 2010). Beyond the effects of categorical perception, transitions along nonhuman–human continua could be disturbing because they undermine the separation between what we identify as us (e.g., human, person) and what we identify as *not* us (e.g., 3D model, robot, ovum; MacDorman & Entezari, 2015; MacDorman, Vasudevan, et al., 2009; Ramey, 2005).

1.2.2. Cognitive dissonance

The negative emotional appraisal of the uncanny valley has been identified with psychological discomfort caused by a conflict between the belief that an entity is human and the belief that the same entity is not human (Hanson et al., 2005; MacDorman, Green, et al., 2009; MacDorman, Vasudevan, et al., 2009; Tondu & Bardou, 2011). If *nonhuman* and *human* are conceived as distinct and mutually exclusive categories, entities whose appearance gradually transitions from nonhuman to human, as in Mori's (1970/2012) graph, must cross a category boundary. An entity crossing the boundary could at once elicit two mutually exclusive concepts (Moore, 2012)—or even oscillate between them as its appearance changes. Repeated nonconscious elicitation and conscious suppression of the concept *human* could interfere with empathy (Misselhorn, 2009).

1.2.3. Categorization difficulty

Another explanation of the uncanny valley effect is that difficulty in categorizing ambiguous entities results in the formation of negative impressions (Yamada et al., 2013). Thus, categorization difficulty predicts that the most ambiguous representations are perceived as the least likeable. Categorization difficulty (i.e., low processing fluency) is operationalized as longer response times during a categorization task.

1.2.4. Limited investigation of category uncertainty

Although the categorical perception of entities lying on a human likeness or animacy continuum has been established (e.g., Cheetham et al., 2011; Looser & Wheatley, 2010), the effect of categorical perception on the viewer's emotional appraisal of an entity was examined only recently (Yamada et al., 2013). In Yamada and colleagues' study, intermediate morphs between a real, hand-drawn, and stuffed-toy human face elicited the longest categorization latency and the lowest ratings of likeability.

The study, however, is not without limitations. First, it does not directly examine the uncanny valley in the domains where it is typically identified: humanoid robotics and 3D computer animation. Second, the study does not rule out a potential extraneous cause of negative emotional appraisals of ambiguous representations: morphing artifacts from feature misalignment (cf. in Fig. 2 of Yamada et al., 2013, one face has two noses). Third, the findings do not indicate whether the faces were uncanny, because the only dependent variable measured was likeability. Fourth, operationalizing the *y*-axis of Mori's graph as likeability could have confounded it with the *x*-axis of human likeness because these measures are highly correlated (e.g., $r = .73$, $p < .001$, in Ho & MacDorman, 2010).

Cheetham et al. (2014) were unable to find support for the categorization difficulty theory. Unfortunately, the only dependent variable measured was subjective familiarity.

1.3. Realism inconsistency theory

To explain the uncanny valley effect, we have developed an alternative theory to category uncertainty—realism inconsistency. Realism inconsistency theory predicts that features at inconsistent levels of realism in an anthropomorphic entity cause perceptual processes in viewers to make conflicting inferences regarding whether the entity is real. Such inconsistency could violate neurocognitive expectancies, resulting in large feedback error signals (Friston, 2010; Rao & Ballard, 1999; Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012). Prediction error could lead to a negative emotional appraisal and avoidance behavior (Cheetham et al., 2011; MacDorman & Ishiguro, 2006a, 2006b). Prior research has found inconsistent realism in an entity's features increases reported eeriness (e.g., in eyes and skin, MacDorman, Green, et al., 2009; MacDorman, Vasudevan, et al., 2009; or in voice and appearance, Mitchell et al., 2011).