Cognition 146 (2016) 22–32

Contents lists available at ScienceDirect

Cognition

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

Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley

Maya B. Mathur [a,*], David B. Reichling [b]

[a] Quantitative Sciences Unit, Stanford University, United States
[b] Department of Oral & Maxillofacial Surgery, University of California, San Francisco, United States

ARTICLE INFO

Article history:
Received 4 July 2015
Revised 6 September 2015
Accepted 7 September 2015

Keywords:
Human–robot interaction
Social interaction
Game theory
Facial appearance
Trust
Social decision-making

ABSTRACT

Android robots are entering human social life. However, human–robot interactions may be complicated by a hypothetical Uncanny Valley (UV) in which imperfect human-likeness provokes dislike. Previous investigations using unnaturally blended images reported inconsistent UV effects. We demonstrate an UV in subjects' explicit ratings of likability for a large, objectively chosen sample of 80 real-world robot faces and a complementary controlled set of edited faces. An "investment game" showed that the UV penetrated even more deeply to influence subjects' implicit decisions concerning robots' social trustworthiness, and that these fundamental social decisions depend on subtle cues of facial expression that are also used to judge humans. Preliminary evidence suggests category confusion may occur in the UV but does not mediate the likability effect. These findings suggest that while classic elements of human social psychology govern human–robot social interaction, robust UV effects pose a formidable android-specific problem.

1. Introduction

Robots are no longer merely features of our technological environment, but are beginning to penetrate our social sphere (Breazeal, 2003; Fong, Nourbakhsh, & Dautenhahn, 2003; Zhao, 2006), and people who interact with robots are increasingly unlikely to be technically trained experts and thus more likely to use casual intuitive approaches to the interaction. Unexpectedly negative reactions to imperfectly human robots have become a major problem in the design of socially interactive robots. This phenomenon (Fig. 1), termed the "Uncanny Valley" (Mori, 1970), has dominated discussion of human reactions to anthropomorphic robots in both popular culture and research literature. Despite its prominence, the existence of an Uncanny Valley (UV) is controversial (Burleigh, Schoenherr, & Lacroix, 2013; Hanson, 2006; Katsyri, Forger, Makarainen, & Takala, 2015; MacDorman, Green, Ho, & Koch, 2009), and a recent systematic review concluded that "empirical evidence for the uncanny valley hypothesis is still ambiguous if not non-existent" (Katsyri et al., 2015). Most studies attempting to address the issue have employed progressively morphed blends of human and robot faces, in which two face images are digitally overlaid with varying degrees of opacity, in some cases enhanced by warping of features in intermediate images. This method introduces unnatural distortions, such as semi-transparent or bent facial features, that would be most prominent in the more highly processed images in the midrange of a morphed face series. This could potentially create an UV-like artifact in that region (Katsyri et al., 2015).

The present study was designed to determine if human reactions to android robots truly exhibit an UV effect, and if so, to determine the degree to which it actually influences humans' willingness to trust a robot as a social partner. Experiment 1 examined human reactions to a large, objectively chosen sample of real-world android robots using subjects' explicit judgments of the mechano-humanness and likability of each face. Next, to determine whether the influence of a potential UV actually penetrates humans' implicit social decision-making, we employed game-theory methodology to measure subjects' practical inferences (as measured by real financial risk-taking) concerning the trustworthiness of each robot. An exploratory analysis tested the theory that UV effects arise from perceptual category confusion.

In contrast to the large, heterogeneous population of wild-type robots (with variable facial expressions, positions, and background settings) of Experiment 1, Experiment 2 took a complementary approach: we used a precisely controlled series of 6 digitally composed robot faces with constant morphometry to assess social responses to a single face configuration in its controlled progression from mechanical to human. In addition, this control over the

* Corresponding author at: Quantitative Sciences Unit, 1070 Arastradero Rd, Palo Alto, CA, United States.
E-mail address: mmathur@stanford.edu (M.B. Mathur).

http://dx.doi.org/10.1016/j.cognition.2015.09.008
0010-0277/

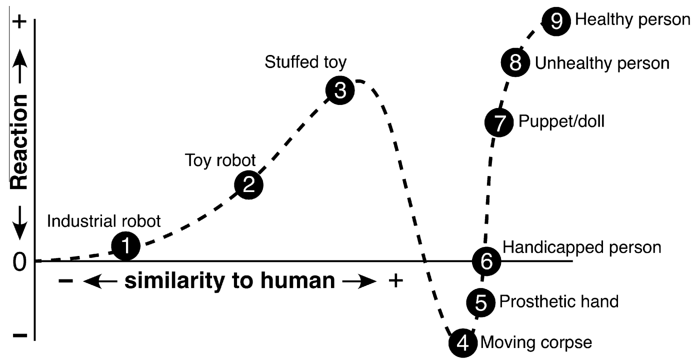

Fig. 1. The Uncanny Valley conjecture, adapted from Mori (1970).

face images allowed us to measure the effect on trust-motivated behavior of a subtle change in facial expression.

2. Experiment 1: wild-type robots

2.1. Experiment 1A: quantifying the mechano-humanness spectrum

2.1.1. Methods

All protocols received IRB approval. Experiments 1A–1C used the same sample of 80 real-world robot faces (Fig. 2) that embodied the myriad design choices made by actual robot designers, choices that may be subtle and unexpected and may vary depending on whether the designer's intention is to build more mechanical versus more human-like robots. The size of the sample and its diversity in mechano-humanness enabled a fine-grained statistical analysis of the effect of mechano-humanness on human social perceptions. To reduce bias in selecting the robots or their manner of presentation (expressions, poses, viewing angles, background settings, etc.), we conducted a systematic search using specific inclusion and exclusion criteria. We performed four Google image searches on a single day using the following sets of search terms: "robot face," "interactive robot," "human robot," and "robot."

Inclusion criteria were:

1. Full face is shown from top of head to chin.
2. Face is shown in frontal to 3/4 aspect (both eyes visible).
3. The robot is intended to interact socially with humans.
4. The robot has actually been built.
5. The robot is capable of physical movement (e.g., not a sculpture or purely CGI representation that lacks a three-dimensional body structure).
6. The robot is shown as it is meant to interact with users (e.g., not missing any hair, facial parts, skin, or clothing, if these are intended).
7. The robot represents an android that is plausibly capable of playing the wagering game (e.g., not a baby or an animal).
8. The resolution of the original image (or an exact copy when one could be located) is sufficient to yield a final cropped image at 100 d.p.i. and 3 in. tall.

Exclusion criteria were:

1. The robot represents a well-known character or a famous person (e.g., Einstein).
2. The image includes other faces or human body parts that would appear in the final cropped image.
3. Objects or text overlap the face.
4. The robot is marketed as a toy.

When the search returned multiple images of a particular robot, we accepted only the first image encountered; if an image failed only graphical criteria, we accepted the next graphically adequate image of the same robot. We accepted the first 80 face images satisfying inclusion criteria and cropped them to include top of head to bottom of chin (or when those features were missing, images were similarly framed in approximate proportion to the features).

For Experiments 1A–1C, we sampled subjects via Amazon Mechanical Turk, a crowdsourcing platform allowing workers to complete brief online tasks in exchange for pay. The task title and description were vague to minimize sampling bias and demand characteristics. We sampled United States workers with excellent performance history (>95% of previous online tasks "approved" as high-quality by requester). By contractual agreement with Amazon, workers must be at least 18 years old. Mechanical Turk workers tend to be somewhat younger, more educated, and lower-income than the US general population, but are demographically more representative than typical university-based research samples (Paolacci, Chandler, & Ipeirotis, 2010). Studies performed on Mechanical Turk can yield high-quality data, minimize experimental biases, and successfully replicate the results of behavioral studies performed on traditional samples (Paolacci et al., 2010). No subjects were duplicated across any of the experiments to avoid effects of previous exposure to the stimuli.

The purpose of Experiment 1A was to determine (1) the degree to which each robot face is perceived as exhibiting human and mechanical properties, (2) the extent to which human-resemblance and mechanical-resemblance behave as a unidimensional property, and (3) secondarily, the perceived valence and magnitude of emotion displayed by each face, which might strongly influence and confound social responses in subsequent experiments (Scharlemann, Eckel, Kacelnik, & Wilson, 2001).

In an online questionnaire, subjects first viewed a page of thumbnails of all 80 face stimuli (similar to Fig. 2 but with faces arbitrarily positioned with respect to mechano-humanness) to give them a sense of the range of the faces they would encounter. Subjects then viewed and rated each of the 80 robot faces one at a time, with the order of faces randomized for each subject. The rating scale was a continuous visual analog scale (VAS) without graduations, which can provide more precise and psychometrically valid ratings than a Likert-type ordinal scale (Reips & Funke, 2008). The subjects controlled how long they viewed each face with no time limit. If individual subjects had rated both the mechanical- and human-resemblance of a face, they might have assumed some relationship between the two properties to which their ratings should conform (e.g., sum to 100). Therefore, in