Neuropsychologia 114 (2018) 181–185

Contents lists available at ScienceDirect

Neuropsychologia

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

Uncanny valley as a window into predictive processing in the social brain

Burcu A. Urgen[a,b,*], Marta Kutas[a,c], Ayse P. Saygin[a,c]

[a] Department of Cognitive Science, UC San Diego, 9500 Gilman Drive, 92093 La Jolla, CA, United States
[b] Department of Medicine and Surgery, University of Parma, Via Volturno 39, 43125 Parma, Italy
[c] Neurosciences Program, UC San Diego, 9500 Gilman Drive, 92093 La Jolla, CA, United States

ARTICLE INFO

Keywords:
Uncanny valley
Predictive processing
N400
Action perception
Social neuroscience

ABSTRACT

Uncanny valley refers to humans' negative reaction to almost-but-not-quite-human agents. Theoretical work proposes prediction violation as an explanation for uncanny valley but no empirical work has directly tested it. Here, we provide evidence that supports this theory using event-related brain potential recordings from the human scalp. Human subjects were presented images and videos of three agents as EEG was recorded: a real human, a mechanical robot, and a realistic robot in between. The real human and the mechanical robot had congruent appearance and motion whereas the realistic robot had incongruent appearance and motion. We hypothesize that the appearance of the agent would provide a context to predict her movement, and accordingly the perception of the realistic robot would elicit an N400 effect indicating the violation of predictions, whereas the human and the mechanical robot would not. Our data confirmed this hypothesis suggesting that uncanny valley could be explained by violation of one's predictions about human norms when encountered with realistic but artificial human forms. Importantly, our results implicate that the mechanisms underlying perception of other individuals in our environment are predictive in nature.

1. Introduction

Our social milieu has changed tremendously in recent years, exposing us to social partners that are dramatically different from those the human brain has evolved with over many generations. Specifically, from guiding students in learning math and science, to helping children with autism and stroke survivors in their exercises, artificial human forms such as robots are rapidly becoming participants in our lives. The introduction of such artificial forms into our lives has in turn allowed us to study the fundamentals of human social cognition, similar to the use of artificial stimuli to learn about the fundamentals of human perception (Gregory, 1980; Rust and Movshon, 2005).

Uncanny valley is a phenomenon that refers to humans' response to artificial human forms, which possess almost human-like characteristics. In describing the phenomenon, Mori (1970), who introduced the term, proposes that the relationship between humanlikeness and humans' response to non-human agents is not a linear one. According to his framework, the increasing humanlikeness of an agent elicits positive responses from humans only up to a certain point, where increasing humanlikeness begins to elicit negative responses, thereby forming a deep valley (Fig. 1). Furthermore, it has been suggested that if the agent is moving, the responses will be more pronounced compared to the static form of the agent. Behavioral studies with humans have provided empirical evidence for the hypothetical curve in Fig. 1 (MacDorman et al., 2009; Thompson et al., 2011; Matsuda et al., 2012; Poliakoff et al., 2013; Cheetham et al., 2013; Piwek et al., 2014; Macdorman and Chattopadhyay, 2016), and studies with non-human primates suggest that it has evolutionary origins (Steckenfinger and Ghazanfar, 2009).

There are several theories that attempt to explain uncanny valley including disease or threat avoidance and mate selection (MacDorman et al., 2009) but these theories lack the potential for scientific testability and are short on providing a mechanistic account of the phenomenon. One other hypothetical mechanism is Bayesian estimation or predictive coding, which is linked to a more general description of neural computational properties of the brain (Rao and Ballard, 1999; Friston, 2010; Moore, 2012), and therefore is a scientifically testable framework. According to predictive coding, the uncanny valley is related to violation of expectations in neural computing when the brain encounters almost-but-not-quite-human agents. A growing body of work has associated Mori's hypothetical curve to the processing of conflicting perceptual or cognitive cues, in which the stimuli are compatible with the elicited expectations or are in violation of them (Ho and MacDorman, 2008; Yamamoto et al., 2009; Mitchell et al., 2011; Cheetham et al., 2011; Saygin et al., 2012; Nie et al., 2012; Cheetham et al., 2013).

Here, we tested the predictive coding theory and its application in action perception (Kilner et al., 2007; Friston, 2010) as an underlying

* Corresponding author at: Department of Medicine and Surgery, University of Parma, Via Volturno 39, 43125 Parma, Italy.
E-mail address: burgen@ucsd.edu (B.A. Urgen).

https://doi.org/10.1016/j.neuropsychologia.2018.04.027
Received 29 September 2017; Received in revised form 23 April 2018; Accepted 24 April 2018
Available online 25 April 2018

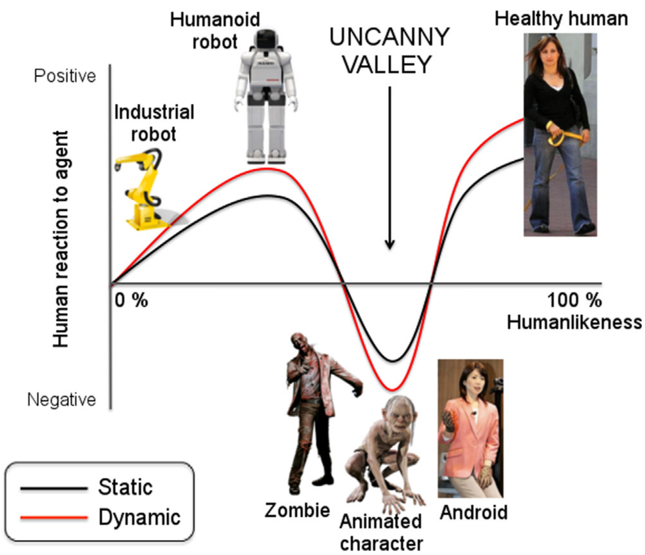

Fig. 1. Hypothetical curves that depict the uncanny valley effect for static and moving agents in varying levels of humanlikeness.

mechanism for uncanny valley. Accordingly, we hypothesize that upon exposure to a human-like form, our brains predict human-like behavior, specifically human-like (biological) movement based on life-long experiences with conspecifics. Uncanny valley occurs when those predictions are not met, such as when faced with agents having human-like forms but non-human-like movements, a hypothesis that has been postulated by Saygin et al. (2012) previously. No empirical work to date has directly tested this theory of prediction violation.

In the present study, we used well-controlled stimuli, which did and did not violate appearance-motion predictions, together with electroencephalography (EEG) and a remarkable biomarker of human information processing, the event-related brain potential (ERP) N400 component to directly test this theory. N400 is the human brain's response to any meaningful stimulus. It is a negative-going event-related brain potential, which peaks around 400 ms after stimulus onset and is maximal in fronto-central regions of the human scalp to pictorial stimuli (Kutas and Federmeier, 2011). Its amplitude is relatively greater for items that violate one's predictions than for items that do not. Thus, it has been linked to the pre-activation of semantic knowledge during comprehension of meaningful stimuli including meaningful actions (Kutas and Federmeier, 2011; Amoruso et al., 2013).

We presented agents of varying humanlikeness in static and dynamic forms as EEG was recorded from human subjects. The stimuli consisted of a real human agent with human-like appearance and motion (Human), a realistic robot agent with human-like appearance and non-human-like motion (Android), and a mechanical robot with non-human-like appearance and motion (Robot) (Fig. 2A). In this stimuli set, the real and mechanical agents (Human and Robot) had *congruent* appearance and motion whereas the realistic agent (Android) had *incongruent* appearance and motion. In this setting, the appearance of the agent provides a context for the subsequent perception of the agent and activates world-knowledge (Metusalem et al., 2012) about agents that have that type of appearance. We hypothesized that the realistic agent (Android) would elicit a greater N400 response in dynamic form than the static form as its human-like appearance would lead to the prediction that it would move in a human-like way based on our world-knowledge but when it did not, it would violate that prediction. On the other hand, we hypothesized that the N400 amplitude for the static and dynamic forms would not differ for Human and Robot since both possess appearance-motion congruence (Human looks human-like, moves in a human-like way; Robot looks non-human-like, moves in a non-human way). Such a pattern of activity would provide direct empirical evidence for the prediction violation theory of uncanny valley.

2. Materials and methods

2.1. Participants

Twenty right-handed adults (10 females; mean age = 23.8; SD = 4.8) from the student community at University of California, San Diego participated in the study. They had normal or corrected-to-normal vision, and no history of neurological disorders. Informed consent was obtained in accordance with the university's Human Research Protections Program. Participants were paid $8 per hour or received course credit. One subject's data was excluded due to high noise during EEG recording.

2.2. Stimuli

Stimuli consisted of video clips of actions performed by the humanoid robot Repliee Q2 (in Robotic and Human-like appearance) and by the human ‘master’, after whom Repliee Q2 was modeled (Fig. 2A, also see Saygin et al., 2012 and Urgen et al., 2013 for details about the stimuli). We refer to these agents as the Robot, the Android (realistic robot), and the Human conditions. Note that the former two are in fact the same robot. Repliee Q2 has 42 degrees of freedom and can make face, head and upper body movements. However, the robot's movements did not match the dynamics of biological motion; it is mechanical or “robotic”. The same body movements were videotaped in two appearance conditions. For the Robot condition, Repliee Q2's surface elements were removed to reveal its wiring, metal arms and joints, etc. The silicone ‘skin’ on the hands and face and some of the fine hair around the face could not be removed but was covered. It is important to note that the movement kinematics of the Android condition was identical to that of the Robot. The silicone skin on the hand or face did not affect the movement kinematics for the Android condition since the performed actions largely included arm and upper torso movements rather than fine detailed finger movements of the hand or face, and the skin was only 1.5 mm and tightly attached to the hand or face. For the Human condition, the female adult whose face was molded and used in constructing Repliee Q2 was videotaped performing the same actions. She was asked to watch each of Repliee Q2's actions and perform the same action naturally. All agents were videotaped in the same room with the same background. Video recordings were digitized, converted to grayscale and cropped to 400 × 400 pixels. Videos were clipped such that the motion of the agent began at the first frame of each video.

2.3. Procedure

Since prior knowledge can affect judgments of artificial agents differentially (Saygin and Cicekli, 2002), each participant was given exactly the same introduction to the study and the same exposure to the videos. Before starting EEG recordings, participants were shown each video and told whether each agent was a human or a robot, and the name of the action. Participants went through a practice session before the experiment. EEG was recorded as participants watched the images or video clips of the three agents performing eight different upper body actions (drinking from a cup, examining an object with hand, hand-waving, turning the body, wiping a table, nudging, introducing self, and throwing a piece of paper). The videos were presented in two modes that we call *motion alone* and *still-then-motion*. In the *motion-alone* condition, 2-second videos were presented. In the *still-then-motion* condition, the first frame of the video was presented for 600–1000 ms (with a uniform probability jitter), and then the full video was played. The experiment consisted of 15 blocks. In each block, the eight videos of each agent were presented once in the *motion-alone* condition, and once in the *still-then-motion* condition. Stimuli were presented in a pseudo-randomized order ensuring that a video was not repeated on two consecutive trials. Each participant experienced a different pseudo-randomized stimuli sequence.