A (Reverse) Developmental Approach to Understanding the Uncanny Valley and Designing Socially-Effective Robots

Megan K. Strait
University of Texas Rio Grande Valley
1201 W University Drive
Edinburg, TX 78539
megan.strait@utrgv.edu

Heather L. Urry
Tufts University
490 Boston Avenue
Medford, MA 02155
heather.urry@tufts.edu

Paul Muentener
Tufts University
490 Boston Avenue
Medford, MA 02155
paul.muentener@tufts.edu

ABSTRACT
Towards facilitating their reception by and communication with human interactants, the development of humanlike agents has received particular interest in robotics and related research areas. The interest is loosely based on similarity-attraction theory, which predicts a monotonically increasing relationship between similarity and liking. However, the emergence of increasingly humanlike robots and computer-generated characters has brought to light a competing phenomenon known as the “uncanny valley”. Contrary to the predictions of similarity-attraction theory, uncanny valley theory suggests a more nuanced relationship between similarity and liking in which certain anthropomorphic entities (often those with a highly humanlike appearance) elicit aversion rather than affinity. Yet, despite substantial supporting evidence, the uncanny valley – the nature of its effects, the underlying mechanisms, and even its existence – remains poorly understood. In particular, critics of the theory argue that without knowing how and when the phenomenon emerges (as well as whether its effects change over time), researchers cannot even begin to address the uncanny valley in the design of social agents. Below we summarize the results from a series of studies and proposals for further testing utilizing a developmental approach to understanding the uncanny. In Studies 1-3 (completed), we drew from research in cognitive and emotion psychology to construct a protocol for measuring aversion as a function of human similarity (in adults). In Studies 4-5 (in-progress), we adapted the protocol for a middle-childhood population (ages 5 to 10 years) and report the preliminary findings (Study 4), as well as the proposed replication procedures (Study 5). Finally we discuss how, taken together, the results of this developmental approach can better inform the design of robots intended for social applications.

KEYWORDS
Uncanny valley theory, developmental paradigms for evaluation of CRI, developmentally-appropriate robot design

1 INTRODUCTION
Within fields concerning the design of artificial agents (e.g., robotics, graphics, animation), there is a pervasive assumption that humanlike agents evoke better responding than do more prototypically-machine-like agents. This assumption is evident from the sheer number of effort devoted to engineering humanlike robots (e.g., [8, 24, 26, 27, 31]) to the instantiation of a new field of study – android science – devoted to this topic [14]. Researchers argue that, by capitalizing on traits that are familiar/intuitive to people, such agents offer more natural and effective interactions than their less humanlike counterparts (e.g., [9]).

This perspective is not unfounded. In fact it is rather well-supported by a large (and still growing) empirical base. Equipping an artificial agent with humanlike features facilitates the formation of rapport with, empathic responding towards, and positive appraisals of the agent (e.g., [2, 29, 36]). People attribute humanlike agents better personalities (e.g., [1, 6, 16, 18]) and assign them positive traits, such as being trustworthy, fair, and intelligent (e.g., [12]). Moreover, people report greater comfort in their presence (e.g., [32, 34]). In contrast, more prototypically-mechanical agents are received in a more negative fashion (e.g., [3, 7, 38]) and are assigned negative traits such as being untrustworthy and unintelligent (e.g., [3, 32]). Such observations are broadly consistent with the theory of similarity-attraction, which predicts that people will respond more positively towards agents the more they resemble themselves (e.g., [10, 11, 21]). With the aim of continued improvement to their reception by and communication with human interactants, this broad empirical and theoretical support has fed into the development of increasingly anthropomorphic artificial agents.

The emergence of increasingly humanlike robots, however, has brought to light an unintended consequence – the uncanny valley [25]. The valley refers to the phenomenon wherein highly humanlike (but not prototypically human) entities provoke aversion in people (for a review, see [15]). For example, highly humanlike robots are rated more negatively [20], avoided more frequently [37], and attributed less trustworthiness [22] than their less humanlike counterparts. Moreover, such effects do not appear to be limited to adults, as valley-like effects have been observed in infants [19, 23], children [39], and even other primates [35], suggesting the general phenomenon is relatively pervasive.

Yet, the uncanny valley continues to be a poorly understood and even contentious research topic, due to gaps in the current literature and various empirical inconsistencies. These issues stem, in large part, from challenges inherent to conducting longitudinal empirical HRI studies (in particular, the limited accessibility of robotic platforms that only partially represent the large design space, as well as the feasibility of their longer-term deployment). This has lead researchers to turn to more accessible alternatives, such as the use of short interaction paradigms involving computer-generated stimuli to make inferences about embodied counterparts (e.g., [13, 28]) and careful case studies of only one or a few robotic platforms (e.g., [4, 17, 33, 36]). But the small range of methodologies for investigating the valley, in turn, has lead to conflicting findings. For example, amongst studies utilizing few robots or non-embodied robot stimuli,
there are both many studies which fail to find a valley effect (or find the opposite – more positive responding to the most humanlike stimuli; [4, 17, 28]) as well as many that confirm its existence (e.g., [33, 37]).

More importantly, these methods have thus far failed to address the questions as to when, why, and how do robots fall into the uncanny valley. In particular, critics question how the valley effects observed in short interaction paradigms extends to more ecologically-valid scenarios. Specifically, how the valley effect emerges and whether it is persistent are two longstanding and open questions that impact the interpretation and relevance of the uncanny valley to HRI. For example, some have challenged the significance of the phenomenon, contending that the effects could arise as a result of socio-cultural conditioning (e.g., exposure to negative portrayals of robots in popular media). Relatedly, critics have questioned the extent to which the valley “lasts” (e.g., [5]) – specifically, given the novelty of robotics platforms, some predict the dissipation of the valley via exposure to such agents. Both critiques thus argue against the relevance of the uncanny valley, given the implication that the valley effects are likely to change over time. Both takes also find support in recent empirical findings ([30, 40]), however, the evidence remains limited in both amount and ability to account for/mitigate the valley.

We have thus employed an alternative approach, drawing on methods from cognitive, affective, and developmental psychology towards tackling these longstanding challenges and open questions. Specifically, we have set out to evaluate valley effects as they manifest in adulthood, childhood, and infancy to determine when and how the uncanny valley emerges. In particular, the aforementioned conditioning argument suggests the onset of the valley is limited to late childhood/early adulthood (post-exposure to media such as the Terminator film series).

2 SUMMARY OF METHODS & RESULTS

Across all studies, we employed a picture-viewing protocol to obtain measurement of people’s emotionally-motivated responding towards a large portion of the current design space in humanoid robotics. The set of pictures used depicts human and robotic agents of varying human similarity from low to high to human (e.g., see Figure 1). We measure participants’ explicit evaluations of the agents, as well as their avoidance of them. See [37] for the full methodological details. At present, over 300 people (adults and children) have participated in our study series.

Studies 1-3: The Development and Validation of Valley-Related Protocol w/Adults. In Study 1 ([37]), we developed our picture-viewing protocol for measuring valley effects in adults, finding that participants were averse to even looking at pictures of highly humanlike robots. Studies 2-3 (in submission) served as replications of and extensions to Study 1, demonstrating in particular the same elicitation of avoidance with two different sets of participants. Together, the three studies established a substantial effect on people’s behavior (avoidance), as well as a protocol for investigating the uncanny valley.

Studies 4-5: Extension to Children. To test whether the uncanny valley emerges/exists before adulthood, we adapted our protocol established via Studies 1-3 for children (ages 5 to 10). In particular, we modified the explicit prompts for participant ratings (regarding how eerie/uncanny an agent is) to a scale of liking/disliking, as pretesting with children revealed that they do not yet understand the concept of uncanniness. While we await the completion of data collection, preliminary descriptive statistics cross all ages (grouped by ages 5-6, 7-8, and 9-10) suggest that children mirror adult responding. Specifically, they show a consistent dislike of the highly humanlike robots, reflected by both their explicit dislike picks (see Figure 2) as well as their picks for most likeable. Qualitatively, children often vocalize feelings of disgust (e.g., “ew”, “gross”, etc.) in response to the highly humanlike robots and provided redundant affirmation of their picks (e.g., saying “yeah that one is the worst!” after already picking the agent as least likeable). An extension to Study 4 (Study 5) is now in development to see whether the results of Study 4 replicate and if so, how the valley impacts children’s interactions with robotic agents.

3 DISCUSSION

We have elected to pursue an extensive series of studies that specifically span childhood through adulthood towards understanding the qualitative experience of the uncanny valley and when in the...
developmental timeline it emerges. Studies 1-3 served to establish a protocol with the standard participant demographics utilized in HRI (college-educated, young adults). A careful adaptation of this protocol run with children (Study 4) has indicated that the valley exists long before previously thought. If replicated (Study 5), the results would underscore a significant need for designing around the uncanny valley – from childhood through adulthood.

In addition, this line of research offers further contributions in the form of methodological considerations (in particular, how to adapt methods developed for adult participants to younger populations), as well as a look at the stability/instability of perceptual processes involving robotic agents over development. Specifically, by assessing both children and adult perceptions, we can better decide amongst the potential theoretical explanations for people’s behavior and the implications that follow. For instance, here we have indications that the valley phenomenon does not change qualitatively in its manifestation across development. This suggests that a conditioning mechanism is less plausible than one linked to biological motivations or more intrinsic factors. Moreover, it suggests that the design considerations – from an aesthetic, pro-anthropomorphization perspective – are relatively similar across 5 to approximately 25 years old.

In sum, we aim to make three primary contributions with this line of research. First and foremost, the work is oriented to understanding the uncanny valley and furthering uncanny valley theory. Beyond this however, we demonstrate a benefit of a developmentally-broad participant pool and add discussion items of relevance beyond the uncanny valley, as to how to adapt various metrics between developmental stages. As future avenues for research, we are interested to extend the present work towards infancy to further test potential differences across stages of development.

REFERENCES

[6] Elizabeth Broadbent, Vinayak Kumar, Xingyan Li, John Sollers, Rebecca Q. Stafford, Bruce A. MacDonald, and Daniel M. Wegner. 2013. Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. PloS one 8, 13 (2013), e72589.

Figure 2: Dislike Frequencies from Study 4. Children were asked to pick which – from a set of four agents (one low in similarity, two of highly humanlike appearances, and one human) – they “liked the least/disliked the most”. Shown are the average choice frequency by agent category, wherein highly humanlike robots were liked the least of all and across participants of ages 5 to 10 years old.