

ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ  
FACULTY OF PSYCHOLOGY AND COGNITIVE SCIENCE

Dawid Ratajczyk

**Perceiving Artificial Agents:  
A Multimethod Investigation into  
the Shape and Variables of the  
Uncanny Valley**

*Doctoral thesis*

Supervisor:  
**dr hab. Paweł Łupkowski, prof. UAM**

Assistant supervisor:  
**dr inż. Marcin Jukiewicz**



Poznań 2025



## Acknowledgements

Dziękuję Pawłowi Łupkowskiemu za bycie świetnym promotorem, który popychał mnie do przodu dokładnie w tych obszarach, w których potrzebowałem pomocy, a zostawiał mi wolną rękę tam, gdzie radziłem sobie dobrze.

Dziękuję mojej partnerce, Annie Adamczuk, za sprawdzanie i konfrontowanie mojej wizji rzeczywistości, oraz mojej siostrze, Paulinie Ratajczyk, za potwierdzanie i współdzielenie tej perspektywy. Dziękuję rodzicom i dziadkom za wsparcie i wiarę we mnie od najmłodszych lat.

Dziękuję społeczności związanej z pokojem 110, w szczególności Michałowi Wyrwie, Maciejowi Rasiowi, Aleksandrze Wasielewskiej i Marcinowi Jukiewiczowi, za liczne dyskusje związane i niezwiązane z tematem tej pracy. Dziękuję współautorom artykułów, które składają się na tę rozprawę, za wspólne godziny spędzone w labie. Dziękuję również przyjaciółom, między innymi Konradowi Mroczo, Madzi Wiśniewskiej i Krzysiu Kulisowi, za wkład w poznanie siebie (a Krzysiu także za użyczenie głosu postaci AI w jednym z badań).



# Contents

Abstract . . . . .	7
Abstract in Polish . . . . .	8
<b>1 Introduction</b>	<b>9</b>
1.1 Goals of the dissertation . . . . .	11
1.2 Structure of the Thesis . . . . .	12
<b>2 Theory and background</b>	<b>14</b>
2.1 The uncanny valley effect . . . . .	14
2.2 Independent and dependent variables . . . . .	15
2.3 Explanations of the phenomenon . . . . .	17
2.4 Motivations and Methodological Directions . . . . .	19
<b>3 Synthesis of own research</b>	<b>22</b>
3.1 Hypotheses and research questions . . . . .	22
3.2 Methods of conducted research . . . . .	25
3.3 Results . . . . .	26
3.3.1 Study I: Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments . . . . .	26
3.3.2 Study II: The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen . . . . .	27
3.3.3 Study III: Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality . . . . .	28
3.4 Discussion . . . . .	29
3.4.1 Clarification of the Shape and Variables of the Uncanny Valley Effect . . . . .	30
3.4.2 Identification of Cognitive and Social Moderators of Uncanny Valley Responses . . . . .	31
3.4.3 Examination of the Interaction Between Appearance and Mind Attribution . . . . .	33
3.5 Summary . . . . .	35

<b>References</b>	<b>36</b>
<b>A Appendices</b>	<b>44</b>
A.1 Study I: Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments . . . . .	44
A.2 Study II: The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen . . . . .	62
A.3 Study III: Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality . . . . .	75
A.4 Confirmation of authorship statements . . . . .	120
A.5 Biographical Note . . . . .	126

---

## Abstract

The Uncanny Valley Hypothesis (UVH) proposes that artificial agents that are almost—but not fully—humanlike elicit feelings of eeriness and discomfort. Despite extensive research, the causes, shape, and moderators of the uncanny valley effect remain debated. This doctoral dissertation aims to advance the understanding of the uncanny valley phenomenon by addressing three major research goals: (1) clarifying the shape and variables of the uncanny valley effect; (2) identifying cognitive and social moderators influencing emotional reactions toward artificial agents; and (3) examining the interaction between agents' visual appearance and attributed mind type. The dissertation consists of three empirical studies employing complementary methods: a large-scale natural language analysis of over 220,000 YouTube comments, immersive virtual reality (VR) experiments, and psychophysiological measurements. Results show that general humanlikeness is associated with increasingly negative sentiment, while specific visual subdimensions—particularly facial features—exhibit non-linear relationships with eeriness. Essentialist beliefs about human uniqueness were found to modulate eerie responses, especially for uncanny humanlike agents. Moreover, while appearance primarily influenced eeriness, the congruence between visual appearance and mind attribution affected perceived behavioral realism and likability. These findings demonstrate that the uncanny valley effect emerges from a complex interaction of perceptual, cognitive, and social processes. The dissertation contributes a multidimensional perspective on human-agent interaction and offers theoretical refinements to UVH, with implications for designing more emotionally acceptable and socially coherent artificial agents.

## Abstract in Polish

Hipoteza doliny niesamowitości (ang. *Uncanny Valley Hypothesis, UVH*) zakłada, że sztuczni agenci, którzy są niemal – ale nie całkowicie – podobni do ludzi, wywołują uczucie niepokoju (ang. *eeriness*) i dyskomfortu. Pomimo rosnącej liczby badań, przyczyny, kształt zależności oraz czynniki moderujące ten efekt pozostają przedmiotem dyskusji. Niniejsza rozprawa doktorska ma na celu pogłębienie zrozumienia zjawiska doliny niesamowitości poprzez realizację trzech głównych celów badawczych: (1) określenie kształtu zależności oraz zmiennych związanych z efektem doliny niesamowitości; (2) identyfikację poznawczych i społecznych moderatorów wpływających na reakcje emocjonalne wobec sztucznych agentów; oraz (3) zbadanie interakcji między wizualną formą agentów a przypisywanym im typem umysłu. Niniejsza dysertacja obejmuje trzy badania empiryczne wykorzystujące komplementarne metody: analizę językową ponad 220 000 komentarzy z portalu YouTube, eksperymenty w immersyjnej rzeczywistości wirtualnej oraz pomiary psychofizjologiczne. Wyniki wskazują, że ogólny poziom podobieństwa do człowieka wiąże się ze stopniowym spadkiem pozytywnego sentymentu. Natomiast niektóre konkretne aspekty wyglądu – zwłaszcza cechy twarzy – wykazują nieliniowy związek z odczuwaniem niepokoju. Przekonania esencjalistyczne na temat wyjątkowości natury ludzkiej moderują reakcję niepokoju, szczególnie wobec postaci z doliny niesamowitości. Co więcej, to głównie wygląd postaci wpływa na ocenę niepokoju, natomiast spójność między wyglądem a przypisywanym umysłem kształtuje postrzeganą realistyczność zachowania i lubialność (ang. *likability*) postaci. Wyniki wskazują, że reakcje w dolinie niesamowitości są efektem złożonej interakcji procesów percepcyjnych, poznawczych i społecznych. Rozprawa dostarcza wielowymiarowej perspektywy na interakcje człowiek-agent oraz wnosi teoretyczne doprecyzowania do modelu doliny niesamowitości, co ma znaczenie dla projektowania społecznie akceptowalnych i emocjonalnie dopasowanych sztucznych agentów.

# Chapter 1

## Introduction

The Uncanny Valley Hypothesis (UVH) proposes that artificial characters that appear almost—but not fully—human elicit feelings of eeriness or discomfort in observers (Mori, 1970; Seyama and Nagayama, 2007). As these characters become increasingly humanlike, they initially evoke positive responses, but upon reaching a certain threshold of resemblance, they may suddenly provoke unease and negative emotional reactions. A more detailed overview of UVH is provided in Chapter 2.1.

Although numerous explanations have been proposed for the causes of this effect, there remains no clear consensus among researchers (see e.g., Kätsyri et al., 2015; Diel et al., 2021). The uncanny valley appears to be a multifaceted phenomenon composed of distinct cognitive and affective mechanisms. Theoretical accounts refer to different systems within the spectrum of perceptual, cognitive, and social processing, leading to a fragmentation of research methodologies across studies. Identifying specific variables and moderators most relevant to the uncanny valley could help to refine and consolidate future research efforts (see discussion in Kätsyri et al., 2015).

One group of the explanations focuses on deviations in perceptual processing—particularly involving faces and biological motion. Distorted or atypical faces, even those not associated with robotics, often evoke negative emotional responses (Rotshtein et al., 2001), suggesting that this component may reflect a broader cognitive mechanism. Most widely discussed explanations in the literature include perceptual mismatch and categorical ambiguity. These are associated with early stages of perceptual processing: *categorical ambiguity* arises when a stimulus falls between established categories (e.g., human vs. non-human), while *perceptual mismatch* refers to conflicting visual cues (e.g., a highly realistic skin paired with artificial eyes). Recent research supports the idea that the perceptual uncanny effect of face processing is best explained by deficits in configural processing, i.e., the holistic perception of facial features and

their spatial relationships (Diel and MacDorman, 2021).

In addition to perceptual explanations, another account of the uncanny valley effect draws on Theory of Mind and *mental state attribution*. This perspective suggests that the way observers attribute mental capacities—such as emotions, intentions, or consciousness—to artificial agents can significantly influence feelings of eeriness. Characters that appear nearly human may trigger conflicting intuitions about their capacity for experience and agency, leading to discomfort. Specifically, the attribution of experience (the capacity to feel) and agency (the capacity to act intentionally) plays a central role in shaping perceived mind (Gray and Wegner, 2012). For example, artificial agents that are visually humanlike and believed to possess internal experiences may be perceived as unsettling or uncanny, as their apparent emotional capacity blurs the boundary between human and machine, challenging observers’ intuitions about what it means to be truly sentient. This view supports the notion that perceptual cues interact with cognitive attributions in complex ways, suggesting that the uncanny valley is not solely a perceptual phenomenon but also a social-cognitive one.

Another explanatory perspective emphasizes *culturally and socially grounded fears* surrounding artificial intelligence and robots, including concerns related to autonomy, control, and potential rebellion. These fears are deeply rooted in cultural narratives, as reflected in the etymology of the word *robot*, which originates from the Czech term *robota*, meaning forced labor or drudgery<sup>1</sup>. Such historical associations continue to shape public perceptions of artificial agents (Thellman et al., 2017). From a social-cognitive standpoint, robots are frequently subjects of social categorization processes and they possibly are perceived as members of a distinct social group (Eyssel and Kuchenbrandt, 2012). Furthermore, from an evolutionary psychology perspective, highly humanlike but unfamiliar agents may be processed as members of an outgroup, potentially triggering aversive reactions similar to those once directed toward rival tribes. These mechanisms suggest that uncanny responses may not only stem from perceptual or cognitive conflicts but also from deeper, socially and evolutionarily ingrained heuristics regarding group membership and threat detection.

Beyond its theoretical significance, the uncanny valley effect has clear societal and ethical implications. In social robotics<sup>2</sup>, it is often assumed that people prefer to interact with machines in ways that mirror their interactions with other humans, which has encouraged the design of in-

---

<sup>1</sup><https://www.merriam-webster.com/dictionary/robot> [Accessed: 25.10.2025].

<sup>2</sup>Social robotics refers to a field of research concerned with the design, development, and study of robots capable of engaging in social interactions with humans (Fong et al., 2003).

creasingly humanlike robots and avatars. Such agents are already being considered for emotionally sensitive domains, such as elder care, where acceptance may depend critically on their appearance and the feelings they evoke (Sharkey and Sharkey, 2012; Robinson et al., 2014). At the same time, children and younger users are increasingly exposed to highly realistic virtual characters in games and to AI-generated influencers, avatars, and chatbots on social media (Tinwell, 2014). Understanding how these near-human agents are experienced is therefore important not only for improving human–robot and human–computer interaction, but also for supporting the ethical design of artificial agents that do not unintentionally provoke discomfort or psychological harm.

Although the uncanny valley effect has attracted substantial attention and a range of candidate mechanisms has been proposed, a key gap remains: it is still unclear which factors most reliably shape eerie responses, and how perceptual cues, mind attribution, and social beliefs combine to produce them. This doctoral project addresses this gap, as outlined in the aims below.

## 1.1 Goals of the dissertation

The main aim of this doctoral project is to advance our understanding of the uncanny valley effect by integrating ecological validity, theoretical refinement, and novel methodologies in human–agent interaction studies. Specifically, this project investigates how visual (i.e., appearance of artificial agents), cognitive (i.e., mind attribution), and social belief factors (i.e., essentialist views on human uniqueness) interact to influence emotional responses—particularly eeriness and likability—toward artificial agents. By combining large-scale data analysis from social media and experimental studies in immersive virtual reality (VR) environments, the research seeks to:

1. Clarify the shape and variables of the uncanny valley;
2. Identify cognitive and social moderators, such as beliefs in human uniqueness and knowledge of UVH, that affect user perceptions;
3. Examine the interaction between appearance and mind attribution in a realistic context.

This multi-method approach aims to contribute both theoretically to UVH literature and practically by deepening our understanding of how artificial agents are perceived, thereby enhancing the potential for more constructive and meaningful human–artificial agent interactions. As will be made explicit in Chapter 3, and especially in the Discussion (3.4), each detailed result will be mapped back to the goals stated here.

## 1.2 Structure of the Thesis

This dissertation takes the form of a series of peer-reviewed empirical studies, each contributing to an understanding of the uncanny valley effect from a different methodological or theoretical perspective:

- “Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments”  
(published in *International Journal of Social Robotics*, 2022)
- “The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen”  
(published in *International Journal of Human–Computer Interaction*, 2023)
- “Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality”  
(accepted for publication in *International Journal of Social Robotics* on 8th of October 2025)

The studies correspond to the main goals of the project outlined in Section 1.1 and together offer a multi-method investigation of mechanisms underlying people’s responses to artificial agents. Their order reflects a progression from mapping uncanny-related reactions in everyday language (Study I), through examining the role of individual beliefs in a controlled laboratory setting (Study II), to testing how appearance and mind attribution jointly shape uncanny feelings in immersive interaction (Study III).

In Chapter 2, I provide the theoretical background for the research. It consists of a discussion of the uncanny valley hypothesis, its origins, core concepts, and the variety of explanations proposed in the literature. Special attention is given to the ambiguity surrounding variables used in UVH-related studies (such as familiarity, humanlikeness, realism, eeriness, and likability).

In Chapter 3, I present a synthesis of the research, including the hypotheses, methodological approaches, results, and discussion. The first study (Appendix A.1) investigates how people describe robots in an unscripted way across the spectrum of humanlikeness using large-scale data from YouTube comments. This study applies natural language processing techniques to over 220,000 user comments to assess the emotional language associated with different types of robotic agents. The goal is to clarify how emotional reactions to robots can be measured and to sketch the overall shape of the uncanny valley across a wide range of real-world robots (see Section 2.2). The results show that very humanlike robots are more often described with eerie terms such as “scary” and “creepy”, that

eeriness (but not pleasantness or attractiveness) is systematically related to how humanlike a robot appears, that a valley-shaped pattern emerges specifically for facial features, and that smaller robots tend to elicit more positive, playful reactions than larger ones.

The second study (Appendix A.2) focuses on the role of individual beliefs—specifically, essentialist views on human uniqueness—in modulating responses to artificial agents. In this experiment, participants encountered four characters ranging from clearly robotic to clearly humanlike in a virtual café (in virtual reality or on a standard computer screen). The study examined the influence of display medium, perceived eeriness, and the relationship between immersion and emotional assessments. The results show that robotic characters were generally perceived as more eerie than humanlike ones, with the moderately humanlike character receiving the highest eeriness ratings, and that this pattern was similar in VR and on-screen. At the same time, stronger beliefs in human uniqueness were associated with higher eeriness specifically for the most uncanny character, suggesting that such beliefs can deepen the uncanny valley.

The third study (Appendix A.3) examines the interaction between agent appearance and mind attribution in shaping emotional reactions. Conducted in a fully immersive virtual reality environment, this study employed a  $2 \times 2$  experimental design, manipulating both visual appearance (robot vs. humanlike) and perceived mind type (AI vs. human user). In addition to self-reported measures of eeriness and likability, the study incorporated psychophysiological data (electrodermal activity and heart rate) to capture affective responses. The results indicate that visual appearance is crucial for uncanny feelings: the robot-looking character was perceived as less humanlike and more eerie, whereas the type of mind attributed to it did not, on its own, reliably change eeriness. At the same time, appearance and mind attribution interacted to influence behavioral realism, likability, and autonomic arousal, suggesting that how an agent looks and who it is believed to be jointly shape the overall quality of the interaction.

The dissertation concludes with a discussion of the findings across all three studies, offering theoretical implications, methodological contributions, and limitations of the studies.

## Chapter 2

# Theory and background

### 2.1 The uncanny valley effect

The uncanny valley hypothesis (UVH) was introduced by Japanese robotics professor Masahiro Mori (1970). Mori proposed that as robots, dolls, and puppets become increasingly humanlike in appearance, observers would generally experience increasing feelings of likability—up to a critical threshold. Beyond this point, however, slight imperfections in human resemblance would cause a sharp drop in likability, resulting in discomfort or eeriness. As humanlikeness continues to increase beyond the valley, likability rises again for figures that closely resemble actual humans. This pattern forms the characteristic curve known as the *uncanny valley*, illustrated in Figure 2.1. In his original essay, Mori illustrated this idea with the example of a prosthetic hand: shaking what is first taken to be a living human hand and only later recognized as artificial can produce an especially intense uncanny reaction. He also speculated that movement would amplify such responses, such that an artificial agent in motion would be experienced as even eerier than the same figure at rest.

While Mori forged the uncanny valley concept, he was not the first to recognize the phenomenon the concept describes. At the beginning of the 20th century, Jentsch (1906/2008) observed that “life-size machines that perform complicated tasks” can evoke feelings of uneasiness and uncanniness. He emphasized that the more sophisticated and genuine a mechanical figure appeared, the stronger the uncanny sensation it could elicit. Freud later expanded on the notion of the uncanny by exploring its psychological underpinnings more broadly, focusing on mechanisms such as repression, the return of the repressed, and disturbances in identity (Freud, 1919/2003).

The more modern formulation of these earlier ideas was articulated in proposing a specific relationship between humanlikeness and emotional response, particularly emphasizing increased eeriness toward entities that

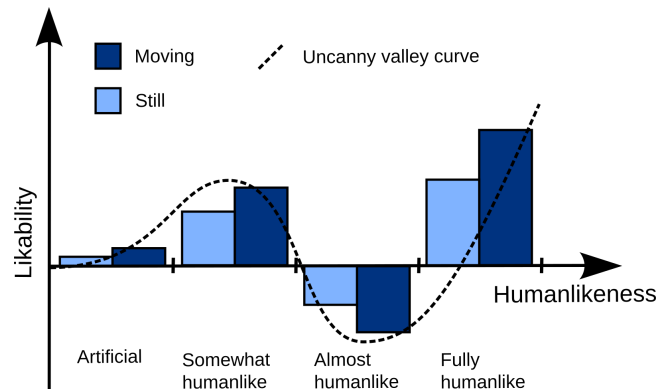


Figure 2.1: Schematic representation of the uncanny valley effect. Adapted from Kätsyri et al. (2015).

are almost—but not fully—human (Mori, 1970). Although Mori originally focused on humanoid robots and puppets, contemporary research has expanded the uncanny valley framework to encompass a broader range of artificial entities, including virtual characters, avatars, computer-animated figures, and conversational agents (Ratajczyk, 2019; Diel et al., 2021). The phenomenon is now recognized as arising not only in physical machines but also in digital environments where humanlike appearance, movement, and behavior are simulated.

Thus, a clearer understanding of the dimensions along which artificial agents vary, and how these dimensions relate to emotional reactions, is essential for advancing research on the uncanny valley phenomenon.

## 2.2 Independent and dependent variables

As research on the uncanny valley effect progressed, it became evident that both the independent variable (typically some form of humanlikeness) and the dependent variable (the affective response to artificial agents) required careful clarification. Although the general idea proposed by Mori was conceptually compelling, he did not conduct any studies to support his proposition. Interest in experimentally studying the uncanny valley began to rise in the mid-2000s (e.g., MacDorman, 2005, 2006), with a rapidly increasing number of empirical publications addressing the topic (Diel et al., 2021). These subsequent empirical studies have revealed considerable ambiguities regarding the precise operationalization of key constructs (Kätsyri et al., 2015).

Originally formulated in Japanese, the translation of the independent variable posed little difficulty: “humanlikeness” was a straightforward and intuitive interpretation. In contrast, the dependent variable, described by

Mori using the term *Shinwakan*, introduced significant challenges. *Shinwakan* is an uncommon Japanese word lacking a direct English equivalent, with proposed translations such as “familiarity,” “affinity,” and others attempting to capture its nuances (see Kätsyri et al., 2015). These translation issues are deeply intertwined with broader ambiguities concerning the emotional responses associated with the uncanny valley (Wang et al., 2015; Kätsyri et al., 2015). As noted by Kätsyri et al. (2015) in their review of the uncanny valley literature, terms such as eeriness, likability, familiarity, and affinity, although frequently used, may correspond to different underlying perceptual or emotional processes. They recommend that “empirical studies would be necessary for resolving which self-report items would be ideal for measuring affinity” (p. 3).

Although several attempts have been made to disambiguate the emotional language used in uncanny valley studies (e.g., Ho et al., 2008; Ho and MacDorman, 2010), the results remain inconclusive. As a consequence, studies variously interpret the dependent variable as perceived warmth (e.g., Kim et al., 2019), eeriness, pleasantness, creepiness (e.g., Kätsyri et al., 2019), likability (e.g., Mathur and Reichling, 2016), acceptance (e.g., Tay et al., 2018), or comfort level (e.g., Łupkowski et al., 2019; Ratajczyk et al., 2019). This lack of terminological consistency complicates comparisons across studies and may blur important distinctions between different types of emotional reactions to artificial agents. In this dissertation, due to this conceptual and terminological heterogeneity, I have decided to use the term *uncanny feeling* as an umbrella term referring to the cluster of negative affective reactions typically associated with the uncanny valley phenomenon—such as eeriness, creepiness, and discomfort. I believe that such a solution is beneficial for UVH-related discussion, as it makes it possible to refer to this family of related negative reactions in a concise way, without assuming that all studies measure one precisely defined emotional construct.

While the concept of humanlikeness, as the independent variable, initially appeared less problematic, it also introduces complexities upon closer examination. Humanlikeness generally refers to the extent to which an artificial agent resembles a human being. However, in studies focusing on computer-animated characters, researchers often substitute or conflate humanlikeness with realism (e.g., Chattopadhyay and MacDorman, 2016; Schindler et al., 2017). The distinction between realism and humanlikeness is not always clear. Some authors treat the two constructs separately (e.g., Phillips et al., 2018), while others use them interchangeably (e.g., Cheetham et al., 2011). Broadly speaking, humanlikeness encompasses the possession of human-like attributes—whether physical, behavioral, or mental—whereas realism pertains more specifically to the fidelity and detail of physical appearance, particularly in visual media (Złotowski et al.,

2014; Diel et al., 2021; Miller, 2021).

Moreover, humanlikeness itself is a multidimensional concept. Scholars have identified three primary domains through which similarity to humans can be assessed: physical (visual), behavioral, and mental attributes (Diel et al., 2021). However, conflating these domains in experimental settings may lead to result misinterpretation, as perceptions of an agent’s physical appearance can influence inferences about its mental capacities and vice versa, a phenomenon often related to the so-called halo effect (e.g., Forgas and Laham, 2016).

Previous research suggests that human perception of artificial agents’ humanlikeness may not be gradual but categorical (Kätsyri et al., 2015). However, the number and nature of these perceptual categories remain unclear, particularly when considered in relation to theories of social categorization such as psychological essentialism (Demoulin et al., 2006). What complicates the picture even more is that factors such as a robot’s displayed emotions can significantly affect perceived humanlikeness (Złotowski et al., 2014).

Emerging evidence further complicates the picture. Some authors propose the existence of two distinct valleys: one for moderately humanlike robots and another for highly humanlike ones (Kim et al., 2022). It remains an open question whether these valleys share the same underlying causes or whether separate mechanisms are responsible.

## 2.3 Explanations of the phenomenon

The UVH remains one of the most prominent issues in human–robot interaction. A study by Mathur and Reichling (2016) demonstrated that the uncanny valley effect not only influences perception but also impacts trust-related social behaviors toward robots. Over the years, several explanations have been proposed to account for the effect, ranging from ultimate evolutionary causes to more proximate perceptual and cognitive mechanisms.

Early theories focused on ultimate causes. The *threat avoidance* hypothesis suggests that eeriness serves as an adaptive response to cues of potential pathogen threats (Rozin and Fallon, 1987; MacDorman et al., 2009; Moosa and Ud-Dean, 2010). The *mortality salience* explanation attributes the uncanny feeling to subconscious reminders of death, triggered by lifeless yet humanlike figures (MacDorman, 2005; MacDorman et al., 2009). Similarly, the *evolutionary aesthetics* hypothesis proposes that unattractiveness—manifested through atypical skin textures, asymmetries, or lack of vitality—provokes negative reactions by signaling genetic unsuitability (Thornhill and Gangestad, 1999; Hanson, 2006; Conway et al., 2007).

While these ultimate explanations are difficult to empirically test due to their abstract nature (Kätsyri et al., 2015), more recent research has focused on proximate causes. Three main proximate explanations have been discussed: categorical ambiguity, perceptual mismatch, and configural processing.

The *categorical ambiguity* hypothesis posits that eeriness arises when an artificial agent cannot be easily categorized as either human or non-human (Kätsyri et al., 2015). Initial studies using morphing techniques between human and robotic faces appeared to support this idea (e.g., Yamada et al., 2013; Burleigh et al., 2013), though other results have been mixed or contradictory (Cheetham et al., 2014; MacDorman and Chattopadhyay, 2016). Moreover, newer findings suggest that categorical ambiguity alone is insufficient to explain the uncanny valley (Kätsyri et al., 2015; Diel and MacDorman, 2021).

The *perceptual mismatch* explanation attributes negative reactions to inconsistencies among visual features—for instance, realistic skin combined with unnatural eyes (Pollick, 2009; Kätsyri et al., 2015; Seyama and Nagayama, 2007). According to this view, the violation of coherent appearance cues disrupts expectations and triggers discomfort.

Building on this, recent work highlights the importance of *configural processing*. Configural processing refers to the holistic perception of faces, where spatial relationships between facial features are critical. A well-known example is the “Thatcher effect,” in which inverted faces with altered internal features (such as upside-down eyes and mouth) appear grotesque when upright but not when inverted (Thompson, 1980). This demonstrates that even small disruptions in expected feature configurations can produce strong perceptual anomalies. Research by Diel and MacDorman (2021) suggests that similar disruptions in configural information may underlie the uncanny valley effect: when facial or bodily features deviate from normal human proportions or spatial relationships, observers may experience eeriness as a result of disrupted face processing mechanisms. The researchers demonstrated that configural disruptions in facial perception better predict eeriness than categorization difficulties or perceptual mismatch, challenging the centrality of these as main causes.

Beyond perceptual factors, cognitive explanations involving mind perception have also gained attention. The *uncanny valley of mind* hypothesis builds on Theory of Mind framework, proposing that eeriness arises when observers attribute inconsistent mental states to near-human agents (Chaminade et al., 2007; Krach et al., 2008). A more explicit model was formulated by Gray and Wegner (2012), who distinguished between two core dimensions of mind perception: *agency* (the capacity to act and exert self-control) and *experience* (the capacity to feel and sense). They showed that beliefs about an agent’s ability to experience emotions, in

particular, modulate uncanny feelings. They argue that the uncanny valley emerges not simply from an agent’s humanlike appearance, but from the perception that such agents possess the capacity to experience emotions—an attribution that violates deep-rooted expectations.

Subsequent studies provided further support. Stein and Ohler (2017) demonstrated that participants judged virtual characters as eerier when they believed the agents were controlled by AI rather than by humans, even though visual appearance remained constant. This suggests that mind attribution can significantly alter emotional responses, independently of physical realism. Similarly, Banks (2020) found that humanlike morphology facilitates the attribution of mental states to robots, influencing emotional judgments through heuristics triggered by anthropomorphic cues.

In summary, contemporary research increasingly supports a multifactorial view of the uncanny valley, in which perceptual, cognitive, and social mechanisms interact to produce feelings of eeriness and discomfort. Perceptual disruptions, especially in specialized face-processing networks, appear critical in eliciting eeriness, while cognitive processes related to mind attribution—particularly attributions of experience—modulate and intensify these emotional reactions. However, the precise relationship between visual humanlikeness, mind attribution, and emotional response remains an open question. Future research should address whether thresholds in humanlikeness directly influence experience attribution, or whether these dimensions operate independently.

## 2.4 Motivations and Methodological Directions

Given the complexity of the uncanny valley effect and the mixed empirical findings reviewed above, it seems important not only to refine theoretical explanations but also to employ methods that balance ecological validity with experimental control.

Historically, two dominant methodological approaches have been used to study the uncanny valley: presenting 2D on-screen stimuli such as images (e.g., Seyama and Nagayama, 2007; Cheetham et al., 2011; Mathur and Reichling, 2016; Palomäki et al., 2018) or video recordings (e.g., Piwek et al., 2014; Kätsyri et al., 2017), and conducting direct interactions with real-world robots (e.g., Bartneck et al., 2009; Yam et al., 2021; Thepsonthorn et al., 2021). Each method has its limitations. On-screen studies tend to isolate visual aspects of stimuli and may not capture the complex dynamics of social interaction. Real-robot studies, while offering greater ecological validity, are comparatively rare, costly, and difficult to control experimentally (Bartneck et al., 2009).

VR offers a promising bridge between these two approaches. It enables

the creation of immersive environments in which participants can experience a sense of presence and social proximity with artificial agents, while still allowing researchers to exert high experimental control. Thus, VR makes it possible to systematically manipulate agents' appearance and behavior, combining the realism of live interaction with the methodological rigor of laboratory research. The widespread manifestation of the uncanny valley effect—across robots, virtual humans, wax figures, and avatars—and the difficulty of reproducing results across studies suggest that the effect is more complex than originally proposed and cannot be explained solely through visual realism. VR-based studies, by enhancing presence—that is, the subjective feeling of “being there” and interacting within the virtual environment rather than merely observing it—offer new opportunities for more precise and ecologically valid testing of theoretical explanations. As suggested by Cheetham et al. (2011), the level of presence itself may be a key modulator of uncanny responses. In this dissertation, VR is therefore used in Studies II and III to investigate how humanlikeness, presence, and mind attribution shape emotional responses toward artificial agents.

To complement laboratory-based approaches, this dissertation also employs large-scale natural language processing (NLP) of YouTube comments as an ecologically valid window into everyday talk about robots and virtual characters. Unlike lab-based self-reports, social-media discourse captures spontaneous, unscripted affective language produced in diverse, naturalistic contexts and at scale, enabling robust estimation across the full spectrum of robots' humanlikeness. Applying NLP to YouTube comments makes it possible to identify recurring affective themes—such as eeriness, pleasantness, and attractiveness—without the constraints of self-report measures or the limited sample sizes typical of laboratory studies. This forms the basis of Study I, which maps how uncanny-related reactions are expressed in everyday language and informs the design and interpretation of the subsequent VR experiments.

Finally, this dissertation adopts a more integrative perspective by investigating the relationship between visual humanlikeness and mind attribution, as formulated in the *uncanny valley of mind* hypothesis. Previous studies have suggested that perceptions of an agent's capacity for emotion and consciousness—its *experience*—play a critical role in uncanny responses (Gray and Wegner, 2012; Stein and Ohler, 2017), and some of this work has already been conducted in immersive settings (e.g., Stein and Ohler, 2017). However, the joint impact of visual humanlikeness and mind attribution has not yet been systematically tested in immersive environments. Study III addresses this gap by combining manipulations of appearance and mind type in VR with self-report and psychophysiological measures, while Study II examines how individual beliefs about human

uniqueness and the medium of presentation modulate eeriness. Taken together, these methodological choices are intended to provide converging evidence on how visual, cognitive, and social factors jointly contribute to the uncanny valley effect.

## Chapter 3

# Synthesis of own research

This chapter presents a synthesis of the empirical studies conducted for the needs of the doctoral project. The aim of the research was to deepen our understanding of the uncanny valley effect by examining how visual appearance, mind attribution, and individual beliefs interact to shape emotional responses toward artificial agents.

Each of the three studies included in the dissertation address different dimensions of this phenomenon, employing diverse methodologies: large-scale analysis of natural language data, immersive virtual reality experiments, and psychophysiological measurements. Each study is tied to specific research questions derived from the theoretical framework discussed in the previous chapter.

This chapter begins by outlining the main hypotheses that guided the empirical work. It then presents the methods used in each study, followed by a summary of results. The discussion section integrates the findings across studies and reflects on their broader implications. Finally, a brief summary concludes the chapter.

Together, these studies form a cohesive and complementary body of work, offering a multi-layered perspective on the uncanny valley and contributing to theoretical, methodological, and applied advancements in the field of human-agent interaction.

### 3.1 Hypotheses and research questions

The section summarizes the main hypotheses and research questions derived for each study.

### **Study I: Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments**

In this study, I analyzed 224,544 YouTube comments posted under 1,515 videos featuring 33 real-life robots that varied systematically in their degree of humanlikeness. The aim was to clarify the structure of emotional dimensions related to the uncanny valley effect using large-scale data. By combining user-generated linguistic data with independently rated measures of robot morphology, the study enabled to research how spontaneous public sentiment reflects key theoretical predictions of UVH.

#### **Hypotheses:**

- **H1:** The shape of the graph representing the relationship between humanlikeness (and its subscales) and sentiment valence toward robots is linear.
- **H2:** Emotional indicators (eeriness, pleasantness, and attractiveness) are equally related to humanlikeness. The relationships between humanlikeness (and its subscales) and emotional indicators are linear.
- **H3:** The size (i.e., height) of robots has an impact on emotions elicited by robots.
  - **H3a:** The smaller a robot, the more it is perceived as playable or related to fun.
  - **H3b:** The bigger a robot, the more it is perceived as threatening and dangerous.

### **Study II: The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen**

This study investigated whether beliefs in human uniqueness and the medium of presentation (VR vs. desktop, i.e., the conditions) affect the perception of eeriness and humanlikeness. The study was conducted in the laboratory with the involvement of 94 participants. Participants encountered four humanoid characters of varying humanlikeness within a simulated café environment, enabling direct comparison of emotional responses between immersive and non-immersive settings.

#### **Research Questions:**

- **RQ1:** Do people assess the humanlikeness in VR and on-screen in the same way?

- **RQ2:** Does perceived eeriness differ for conditions and characters?
- **RQ3:** Is immersion level related to eeriness assessment?
- **RQ4:** Are essentialism beliefs related to eeriness assessments?
- **RQ5:** Is UV knowledge related to different eeriness assessments?

### **Study III: Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality**

This study examined how an agent's appearance (robot vs. human) and the type of attributed mind (AI vs. human user) jointly affect self-reported and physiological measures of eeriness and likability in a VR environment. Conducted in laboratory with 116 participants, it combined self-reports and psychophysiological measures (electrodermal and heart activity) to provide a multidimensional assessment of uncanny feelings. By integrating type of mind attribution with controlled manipulations of agent appearance, the study examined how perceptual and cognitive factors interact to shape the uncanny valley effect in realistic social contexts.

#### **Hypotheses:**

- **H1a:** The robotic appearance of a character evokes stronger feelings of eeriness than a more human appearance.
- **H1b:** The attribution of an AI mind to an agent evokes stronger feelings of eeriness than the attribution of a person's mind.
- **H2a:** The human appearance of a character evokes stronger likability than a robotic appearance.
- **H2b:** The attribution of a person's mind to an agent evokes stronger likability than the attribution of an AI mind.

#### **Research Questions:**

- **RQ1c:** Does the appearance of a character evoke different levels of eeriness when different types of minds are attributed to them?
- **RQ2c:** Does the appearance of a character evoke different levels of likability when different types of minds are attributed to them?
- **RQ3a:** Does the robotic appearance of a character evoke different psychophysiological reactions (EDA, HR) compared to a human appearance?

- **RQ3b:** Does the attribution of an AI mind evoke different psychophysiological reactions (EDA, HR) compared to the attribution of a human mind?
- **RQ4a:** How do the type of attributed mind and the appearance of characters influence humanlikeness?
- **RQ4b:** How do the type of attributed mind and the appearance of characters influence realism?
- **RQ5a:** Are the emotional results mediated by the humanlikeness of characters?
- **RQ5b:** Are the emotional results mediated by the realism of characters?

## 3.2 Methods of conducted research

This doctoral dissertation employs a multi-method research design, combining large-scale natural language analysis, immersive virtual reality experiments, and psychophysiological measurements. Each of the three studies addresses a distinct aspect of the uncanny valley effect.

### Study I: Natural Language Processing of YouTube Comments

The first study involved large-scale data analysis of public discourse on robots using natural language processing techniques. A dataset of 224,544 comments was collected from 1,515 YouTube videos featuring physically existing robots across a wide humanlikeness spectrum. Sentiment analysis was conducted using the AFINN lexicon, and custom indices were developed to identify expressions of eeriness, pleasantness, and attractiveness. Humanlikeness ratings for each robot were obtained from the Anthropomorphic roBOT (ABOT) database<sup>1</sup>, including the overall index and three morphology-focused subscales: Surface Look, Body–Manipulators, and Facial Features. The analysis examined how these emotional indicators related to humanlikeness and other factors such as robot size. This study provided ecologically valid, large-scale insights into public attitudes toward artificial agents.

---

<sup>1</sup><https://www.abotdatabase.info/collection> [Accessed: 26.12.2025]

## **Study II: Beliefs in Human Uniqueness and the Role of Immersion**

The second study investigated the role of individual beliefs and presentation medium (VR vs. desktop) in shaping responses to artificial agents. Participants interacted with four agents of varying degrees of humanlikeness within a virtual café environment, either using a VR headset or a standard computer screen. After each interaction, participants rated the agents on humanlikeness and eeriness scales. Additionally, participants completed a questionnaire assessing beliefs in human nature uniqueness (BHNU) (Pochwatko et al., 2015) and an immersion scale (Jennett et al., 2008), adapted for Polish speakers by (Strojny and Strojny, 2014), enabling the assessment of how essentialist beliefs and experienced immersion relate to affective responses toward agents with different visual characteristics.

## **Study III: Appearance and Mind Attribution in Virtual Reality**

The third study used a  $2 \times 2$  between-subjects experimental design conducted in VR. Participants were randomly assigned to interact with an agent that varied in two dimensions: appearance (robotic vs. human) and mind attribution (AI vs. human user). Participants engaged in a spoken, game-like interaction with virtual agents. Emotional responses were measured through both self-report (eeriness, likability, behavioral and appearance realism, humanlikeness) and physiological recordings, including electrodermal activity (EDA) and photoplethysmography (PPG) for heart rate (HR). The study aimed to assess how appearance and mind attribution interact to shape affective and physiological reactions, and whether these effects are mediated by perceived realism and humanlikeness.

## **3.3 Results**

This section summarizes the key findings from the three empirical studies conducted for the needs of the doctoral project.

### **3.3.1 Study I: Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments**

- **H1 and H2:** Emotional indicators showed distinct patterns of association with humanlikeness. Both sentiment and eeriness exhibited a linear relationship with general humanlikeness; however,

when analyzed through the Facial Features subscale, a nonlinear (valley-shaped) relationship emerged. In contrast, pleasantness and attractiveness demonstrated no consistent relationship with humanlikeness. These findings indicate that emotional dimensions are not equally or uniformly related to humanlikeness, thereby partially contradicting the original hypotheses.

- **H3:** Robot size significantly influenced emotional reactions.
  - **H3a:** Smaller robots were more often associated with positive, playful, or affectionate terms, supporting the idea that small robots are perceived as toys or companions.
  - **H3b:** Contrary to expectations, larger robots were not systematically associated with increased perceptions of threat or danger.

The findings highlighted that spontaneous public emotional reactions toward robots vary across emotional dimensions and depend not only on humanlikeness but also on specific morphological features and contextual interpretations.

### 3.3.2 Study II: The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen

- **RQ1:** Participants assessed humanlikeness similarly across VR and desktop conditions, indicating that medium of presentation had no significant effect on perceived humanlikeness.
- **RQ2:** A moderate uncanny valley effect was observed in eeriness ratings, particularly for the moderately humanlike character, but no significant difference between VR and desktop conditions emerged for eeriness.
- **RQ3:** No significant relationship between immersion and eeriness ratings was found.
- **RQ4:** Essentialist beliefs in human uniqueness were positively associated with eeriness ratings of the most uncanny agent, supporting the idea that individual differences in cognitive belief systems modulate uncanny valley responses.
- **RQ5:** Prior knowledge of the uncanny valley effect did not significantly affect eeriness assessments.

Overall, these results suggest that beliefs about human uniqueness play an important role in shaping emotional responses toward uncanny agents. Individuals who more strongly endorsed essentialist views—that is, the belief in a clear and fundamental distinction between humans and non-humans—were more likely to experience heightened eeriness when interacting with artificial characters. In contrast, neither prior knowledge of UVH nor the type of display medium significantly influenced participants' perceptions of humanlikeness or eeriness, indicating that the uncanny valley response is primarily modulated by cognitive belief systems rather than by familiarity with the concept or by the level of immersion.

### 3.3.3 Study III: Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality

- **H1a and H1b** Robotic appearance significantly increased feelings of eeriness compared to human appearance, however no significant main effect of mind attribution (AI vs. human mind) on eeriness was found.
- **H2a and H2b:** Hypotheses were not confirmed, as either the appearance of the characters or type of mind did not evoke differences in likability.
- **RQ1c and RQ2c:** No significant interaction between appearance and mind attribution on eeriness was found; however, a significant interaction between appearance and mind attribution on likability did emerge.
- **RQ3a:** For analysis of EDA, robotic appearance evokes a smaller Skin Conductance Reaction than a human appearance. Different appearances did not alter HR measure.
- **RQ3b:** Attribution of type of mind influenced HR, which was greater in the user condition.
- **RQ4a:** Only the appearance of characters significantly influenced perceived humanlikeness. Characters with a human appearance were rated as significantly more humanlike than robot-looking characters. The type of attributed mind (AI vs. human user) did not significantly influence perceived humanlikeness, and no interaction between appearance and mind type was found.
- **RQ4b:** The appearance of the characters influenced appearance realism. Human characters are perceived as having a more real-

istic appearance than robot characters. Attributed type of mind, together with character appearance, influenced behavioral realism. Robotic characters with AI minds and human characters with user minds are perceived as having more realistic behavior than robotic characters with user minds and human characters with AI minds.

- **RQ5a:** Mediation analysis showed that humanlikeness mediated the relationship between the appearance (robot vs. human) and perceived eeriness. Robotic appearance decreased humanlikeness, which in turn increased eeriness.
- **RQ5b:** Appearance realism did not mediate the relationship between appearance and eeriness. However, behavioral realism mediated the relationship between the interaction of appearance and mind attribution and likability. Thus, behavioral realism (but not appearance realism) was a significant mediator for emotional responses (specifically, for likability).

The findings underline that both visual and cognitive factors contribute to human–artificial agent interaction, although the appearance of agents appears to play a pivotal role in the uncanny valley effect. Specifically, visual cues related to morphology exerted an effect on eeriness and humanlikeness, whereas cognitive attributions of mind type did not significantly influence these measures. At the same time, interactions between appearance and mind attribution in shaping behavioral realism and likability suggest that perceptual and inferential processes jointly determine the general affect toward artificial agents. These results therefore highlight the primacy of perceptual appearance in eliciting uncanny responses, while indicating that cognitive beliefs modulate social-affective evaluations in more subtle, context-dependent ways.

### 3.4 Discussion

This doctoral project aimed to advance the understanding of the uncanny valley phenomenon by (1) clarifying the shape and variables of the uncanny valley; (2) identifying cognitive and social moderators of emotional responses to artificial agents; and (3) exploring how visual appearance and mind attribution interact in shaping human perceptions in a realistic context. The findings from three complementary studies provide important insights into each of these research aims.

### 3.4.1 Clarification of the Shape and Variables of the Uncanny Valley Effect

The first aim of the dissertation was to clarify the shape of the uncanny valley curve and better define the relevant variables associated with emotional responses toward artificial agents.

The analysis of humanlikeness and sentiment showed a generally linear relationship: as robots become more humanlike, sentiment becomes more negative. However, when examining humanlikeness subscales, a more nuanced pattern emerged. Specifically, the Facial Features subscale displayed a non-linear relationship with eeriness and sentiment. This dimension, reflecting the presence of facial structures such as eyes, mouth, and head, exhibited a sinusoidal pattern in emotional responses: very high scores on Facial Features were associated with heightened eeriness and decreased sentiment, resembling the dip predicted by UVH.

Mori (1970) originally hypothesized a non-linear relationship between humanlikeness and affinity. However, empirical evidence has suggested that affinity increases linearly with humanlikeness (Burleigh et al., 2013; Kätsyri et al., 2015). The present analysis of YouTube comments similarly supports a linear relationship between humanlikeness and sentiment, but in the opposite direction than suggested by Kätsyri et al. (2015): as humanlikeness increased, sentiment became more negative. Also, eeriness increased with increasing humanlikeness, supporting a reverse relationship.

Surface Look (e.g., presence of eyelashes, skin texture, apparel) and Facial Features were identified as critical dimensions for predicting emotional responses. Higher Surface Look scores were associated with more negative sentiment. For eeriness, Surface Look initially increased eeriness ratings up to a certain threshold, beyond which it decreased. In contrast, Facial Features exerted a strong non-linear influence on both sentiment and eeriness, reinforcing the importance of social expectations in responses to humanlike robots, suggesting that curve-type UVH relationship is related to processing of faces.

Mori (1970) speculated that robot movement would exaggerate eeriness; however, simple mechanical movements have not consistently confirmed this hypothesis (Piwek et al., 2014). It is possible that complex, socially meaningful movements—not mere motion per se—are necessary to trigger heightened eeriness. This supports the idea that Theory of Mind processes (Dennett, 1971) contribute to the uncanny valley effect and suggests the need for future research on robot behavior in relation to UVH. The involvement of mentalization in uncanny valley effect was explored in the third study of this dissertation.

Analysis of the shape of the relationship also raises questions about the “recovery” portion of Mori’s original curve. My results, con-

sidering mind attribution, suggest a steeper drop in emotional responses—resembling a cliff rather than a valley (Bartneck et al., 2007). It remains an open question whether the rightmost segment of the UVH plot accurately reflects real emotional responses, especially given the current technological limitations in producing robots that are nearly indistinguishable from humans.

Regression analyses indicated that pleasantness and attractiveness were not related to humanlikeness. Only eeriness was significantly associated with humanlikeness. Furthermore, sentiment expressed in YouTube comments often employed words linked to eeriness when describing highly humanlike robots. Importantly, the findings do not support explanations of the uncanny valley based on attractiveness related to mate selection (Burleigh et al., 2013; MacDorman et al., 2009), as no significant relationship was found between attractiveness ratings and humanlikeness in the present data.

It is noteworthy that some YouTube commenters explicitly referred to the uncanny valley effect, suggesting an awareness of the phenomenon within the broader internet community—whether based on formal knowledge or intuitive understanding. The question of whether awareness of UVH can influence uncanny valley responses was addressed in the second study.

Lastly, a significant relationship between robot height and sentiment, mediated by the perception of playfulness, was observed. Smaller robots were seen as more playful and received more positive evaluations.

### 3.4.2 Identification of Cognitive and Social Moderators of Uncanny Valley Responses

The second goal of the doctoral project was to identify cognitive and social moderators, such as beliefs in human uniqueness and knowledge of the UVH, that affect emotional responses toward artificial agents.

The results confirmed that robotic characters were generally perceived as more eerie than humanlike characters. Notably, the highest eeriness ratings were recorded for the moderately humanlike robot. This finding contrasts with some earlier studies suggesting that mechanical-looking artificial agents are consistently perceived as less eerie than humanlike artificial agents (e.g., Yam et al., 2021). One potential explanation, proposed by Tinwell et al. (2011), is that uncanniness may arise from difficulties in detecting emotions, which in turn hinder predictions about behavior. In the study, the android figure lacked facial expressions, possibly complicating mind attribution and contributing to heightened uncertainty regarding the agent's intentions.

Contrary to previous suggestions that immersion level or medium of presentation might influence uncanny valley responses (Hepperle et al.,

2020), participants in this study assessed the humanlikeness and eeriness of characters similarly in both VR and on-screen conditions. Although the VR condition produced higher reported immersion levels, immersion itself was not significantly related to eeriness ratings. This challenges the assumption that heightened immersion necessarily amplifies the uncanny valley effect, at least in the context of relatively simple social interactions.

One of the most important findings relates to essentialist beliefs about human nature. Essentialism—the tendency to perceive social groups as having inherent, natural, and immutable characteristics (Demoulin et al., 2006)—appears to modulate emotional responses to artificial agents. Participants with stronger beliefs in human nature uniqueness reported higher levels of eeriness toward the moderately humanlike robot. This suggests that cognitive schemas about human identity play a significant role in uncanny valley responses, particularly for agents near the human–nonhuman boundary.

These findings extend previous research linking essentialist beliefs to negative attitudes toward robots (Piçarra, 2014; MacDorman and Entezari, 2015; Złotowski et al., 2017). Importantly, unlike previous studies, the current results show that the relationship between essentialism and eeriness is not uniform across all artificial agents but is specific to the most uncanny characters.

Two implications arise from this result. First, essentialism should be considered as a critical variable in future uncanny valley research. Stronger beliefs in human nature uniqueness may be associated with perceived threats to human identity during interactions with near-human agents. This interpretation aligns with findings that people who highly value social group membership exhibit stronger aversion to autonomous artificial intelligence agents (Stein et al., 2019).

Second, these results invite reconsideration of the traditional category uncertainty explanation of the uncanny valley. Previous research has emphasized perceptual ambiguity—difficulty categorizing agents as either human or non-human—as the primary cause of eeriness (e.g., Burleigh et al., 2013; MacDorman and Chattopadhyay, 2016; Palomäki et al., 2018). Although humanlikeness perception appears categorical (Kätsyri et al., 2015), empirical support for the idea that category ambiguity directly induces negative emotions remains weak. Instead, the present findings suggest that the uncanny valley may relate more to social categorization processes than to perceptual category uncertainty. Specifically, moderately humanlike robots may pose a symbolic threat to human social identity, particularly for individuals with strong essentialist beliefs.

Although the present results do not definitively resolve this theoretical issue, they point toward a promising reinterpretation of the uncanny valley effect: agents located near the boundary between nonhuman and

---

human categories may not simply be perceptually ambiguous but may also threaten the integrity of the human social category. Future research should explicitly investigate the interplay between social categorization, perceived group threat, and emotional responses toward artificial agents.

### 3.4.3 Examination of the Interaction Between Appearance and Mind Attribution

The third goal of this doctoral project was to explore how the appearance of artificial agents and the type of mind attributed to them interact to influence emotional reactions in a realistic, immersive context. The findings of Study III provide valuable insights into how these factors jointly contribute to the uncanny valley effect. The results demonstrate that an agent’s appearance plays a crucial role in shaping perceptions of humanlikeness and feelings of eeriness. Robotic appearance significantly increased eeriness ratings compared to human appearance. However, the type of attributed mind—whether participants believed the agent was controlled by an AI or a human user—did not significantly affect perceived eeriness. Thus, perceptions of an agent’s autonomy or human control did not significantly modulate the uncanny valley effect related to eeriness.

Importantly, an interaction between appearance and type of mind attribution was observed for behavioral realism and likability. AI minds attributed to robot-looking agents and human (user) minds attributed to human-looking agents were rated as more behaviorally realistic and subsequently more likable. In contrast, robot-looking users and human-looking AIs were rated as less realistic and less likable, suggesting that incongruence between appearance and attributed mind can negatively impact social evaluations.

Although appearance predominantly determined feelings of eeriness, the mediation analysis revealed a more nuanced pattern. Eeriness was influenced through perceptions of humanlikeness, whereas the interaction between appearance and mind type influenced likability via behavioral realism. These findings align with previous suggestions that eeriness and likability are distinct dimensions of emotional response in human–robot interaction, affected by different cognitive processes (Złotowski et al., 2015). As indicated by results from Study I and Kim et al. (2022), eeriness appears to be a more direct and reliable measure of the uncanny valley effect.

The findings also indicate that cues related to emotional experience attribution—such as verbal priming before experiment or conversational emotional cues—did not influence perceived eeriness. This suggests that visual features of agents are more critical for eliciting emotional responses in UVH, and contradicts the predictions made by Gray and Wegner

(2012).

The robotic character used in the study, while possessing a humanoid body shape, lacked detailed facial expressions, featuring only minimal indicators such as an LED mouth. Prior research has consistently shown that robots without expressive facial features are perceived as less humanlike, less capable of experiencing emotions, and less desirable social partners (Fink, 2012; Phillips et al., 2018; Broadbent et al., 2013).

The inability to accurately recognize emotions may be a critical driver of uncanny experiences. In the Study III, EDA measures supported these findings: stronger Skin Conductance Responses were observed for human-looking characters, indicating greater physiological arousal and engagement compared to robotic-looking ones. This is consistent with higher emotional engagement typically associated with more expressive agents (Banks et al., 2012). Thus, the lack of expected expressive behavior in robotic-looking agents likely contributed to heightened feelings of eeriness. In contrast, HR was primarily influenced by the type of attributed mind, with higher HR observed when participants believed they were interacting with a human user rather than an AI. These results suggest that appearance primarily modulates affective arousal, whereas cognitive attributions of agency influence physiological engagement, underscoring the complementary but distinct roles of visual and cognitive factors in shaping emotional responses to artificial agents.

Supposedly, expectancy violations further elucidate the observed effects. Participants may have implicitly expected that human characters would exhibit rich emotional behavior, while robot characters would not. When these expectations were not met—particularly when a humanlike appearance was combined with AI attribution—negative emotional responses intensified. Previous research similarly suggests that violations of nonverbal behavior expectations (such as gaze, head nodding, and gestures) can trigger feelings of unease (Thepsoonthorn et al., 2021; Schein and Gray, 2015).

Moreover, the social categorization of agents appears to play a role. Participants generally found robot-looking AIs and human-looking users more likable than robot-looking users and human-looking AIs. This pattern suggests that people expect AI agents to look different from humans, possibly reflecting an automatic categorization of AI as an outgroup (Eyssel and Kuchenbrandt, 2012). When these expectations were violated—such as when an AI was presented with a humanlike appearance—feelings of discomfort and decreased likability happened.

In conclusion, these findings underscore the importance of appearance to the uncanny valley phenomenon, while also highlighting the complex interaction between visual features, mind attribution, and expectancy violations. Although mind type attribution alone did not significantly

---

impact eeriness, its interaction with appearance significantly influenced behavioral realism and likability. These results suggest that emotional self-reports related to eeriness and likability should be treated as separate constructs in human–robot interaction research, each influenced by different cognitive mechanisms.

### 3.5 Summary

Results of the studies discussed above offer complementary perspectives on the complex nature of the uncanny valley effect. Together, they contribute new theoretical, methodological, and potential applications to the field of human–agent interaction.

First, the large-scale analysis of language data revealed that the relationship between humanlikeness and emotional responses is more nuanced than originally proposed. While general sentiment toward robots became increasingly negative with greater humanlikeness, specific features—particularly facial features—elicited a non-linear, valley-shaped pattern consistent with the uncanny valley effect. This highlights the need to differentiate between general appearance attributes and critical social-cognitive cues in explaining uncanny experiences.

Second, the investigation into cognitive and social moderators showed that individual differences in beliefs about human nature uniqueness significantly influence emotional reactions to near-human agents. Stronger essentialist beliefs were associated with heightened eeriness, particularly for moderately humanlike characters. These findings point toward the importance of social categorization and symbolic group identity processes in understanding the uncanny valley, extending existing theories beyond perceptual categorization.

Third, the experimental examination of appearance and mind attribution interactions in immersive VR contexts revealed that visual features dominate emotional responses related to eeriness, while the perceived realism of behavior—shaped by the congruence of appearance and attributed mind—affects likability. These results underscore the necessity of considering multiple, distinct emotional dimensions when assessing human responses to artificial agents.

Overall, the studies demonstrate that uncanny valley effect cannot be fully explained by visual factors alone but is shaped by a complex interaction of visual, cognitive, and social factors. This multidimensional perspective offers a more comprehensive framework for future research and practical design strategies in the development of artificial agents intended for social environments.

The present findings open up several avenues for future work. A natural next step is to extend the focus beyond static appearance and

mind attribution to the dynamics of interaction—how artificial agents move, behave, and occupy different social roles in relation to humans, and how these behaviors align (or fail to align) with people’s expectations of their agency and capacity for experience. Examining how such behavioral cues interact with humanlikeness, mind perception, and individual beliefs may help to further refine the proposed framework and bring it closer to the complexity of real-world human–agent encounters.

## **Funding and Note on AI Assistance**

The author acknowledges the use of OpenAI’s ChatGPT (GPT-5) to assist with language refinement and stylistic editing. All conceptual content, interpretations, and conclusions are the author’s own. Study II and Study III were supported by the National Science Centre, Poland, under the PRELUDIUM Grant no. 2020/37/N/HS6/01372, awarded to the author of this doctoral dissertation.

# Bibliography

- Banks, J. (2020). Theory of mind in social robots: replication of five established human tests. *International Journal of Social Robotics*, 12(2):403–414.
- Banks, S. J., Bellerose, J., Douglas, D., and Jones-Gotman, M. (2012). Bilateral skin conductance responses to emotional faces. *Applied psychophysiology and biofeedback*, 37:145–152.
- Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). Is the uncanny valley an uncanny cliff? In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 368–373. IEEE.
- Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2009). My robotic doppelgänger-A critical look at the uncanny valley. In *RO-MAN 2009 – The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 269–276. IEEE.
- Broadbent, E., Kumar, V., Li, X., Sollers 3rd, J., Stafford, R. Q., MacDonald, B. A., and Wegner, D. M. (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(8):e72589.
- Burleigh, T. J., Schoenherr, J. R., and 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.
- Chaminade, T., Hodgins, J., and Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters’ actions. *Social Cognitive and Affective Neuroscience*, 2(3):206–216.
- Chattopadhyay, D. and MacDorman, K. F. (2016). Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision*, 16(11):7–7.

- Cheetham, M., Suter, P., and Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: more like a “happy valley”. *Frontiers in psychology*, 5:1219.
- Cheetham, M., Suter, P., and Jäncke, L. (2011). The Human Likeness Dimension of the “Uncanny Valley Hypothesis”: Behavioral and Functional MRI Findings. *Frontiers in Human Neuroscience*, 5:126.
- Conway, C., Jones, B., DeBruine, L., and Little, A. (2007). Evidence for adaptive design in human gaze preference. *Proceedings of the Royal Society B: Biological Sciences*, 275(1630):63–69.
- Demoulin, S., Leyens, J.-P., and Yzerbyt, V. (2006). Lay theories of essentialism. *Group Processes & Intergroup Relations*, 9(1):25–42.
- Dennett, D. C. (1971). Intentional systems. *The journal of philosophy*, 68(4):87–106.
- Diel, A. and MacDorman, K. F. (2021). Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision*, 21(4):1–1.
- Diel, A., Weigelt, S., and MacDorman, K. F. (2021). A meta-analysis of the uncanny valley’s independent and dependent variables. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(1):1–33.
- Eyssel, F. and Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, 51(4):724–731.
- Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings 4*, pages 199–208. Springer.
- Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166.
- Forgas, J. P. and Laham, S. M. (2016). Halo effects. In Pohl, R. F., editor, *Cognitive illusions*, pages 276–290. Psychology Press.
- Freud, S. (1919/2003). *The Uncanny*. Penguin Books. Original work published 1919.
- Gray, K. and Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1):125–130.

- Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*, pages 39–42. Vancouver, BC.
- Hepperle, D., Ödell, H., and Wölfel, M. (2020). Differences in the uncanny valley between head-mounted displays and monitors. In *2020 International Conference on Cyberworlds (CW)*, pages 41–48. IEEE.
- Ho, C.-C. and MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518.
- Ho, C.-C., MacDorman, K. F., and Pramono, Z. D. (2008). Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 169–176.
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661.
- Jentsch, E. (1906/1997). On the psychology of the uncanny (1906). *Angelaki: Journal of the Theoretical Humanities*, 2(1):7–16.
- Kätsyri, J., de Gelder, B., and Takala, T. (2019). Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images. *Perception*, 48(10):968–991.
- Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, 6:390.
- Kätsyri, J., Mäkäräinen, M., and Takala, T. (2017). Testing the ‘uncanny valley’ hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli. *International Journal of Human-Computer Studies*, 97:149–161.
- Kim, B., de Visser, E., and Phillips, E. (2022). Two uncanny valleys: Re-evaluating the uncanny valley across the full spectrum of real-world human-like robots. *Computers in Human Behavior*, 135:107340.
- Kim, S. Y., Schmitt, B. H., and Thalmann, N. M. (2019). Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing Letters*, 30(1):1–12.

- Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., and Kircher, T. (2008). Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLOS ONE*, 3(7):1–11.
- Lupkowski, P., Rybka, M., Dziedzic, D., and Włodarczyk, W. (2019). The background context condition for the uncanny valley hypothesis. *International Journal of Social Robotics*, 11(1):25–33.
- MacDorman, K. F. (2005). Mortality salience and the uncanny valley. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005.*, pages 399–405. IEEE.
- MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, pages 26–29.
- MacDorman, K. F. and Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, 146:190–205.
- MacDorman, K. F. and Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interaction Studies*, 16(2):141–172.
- MacDorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, 25(3):695–710.
- Mathur, M. B. and Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146:22–32.
- Miller, A. (2021). Realism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition.
- Moosa, M. M. and Ud-Dean, S. M. (2010). Danger avoidance: An evolutionary explanation of uncanny valley. *Biological Theory*, 5(1):12–14.
- Mori, M. (1970). The uncanny valley. *Energy*, 7(4):33–35.
- Palomäki, J., Kunnari, A., Drosinou, M., Koverola, M., Lehtonen, N., Halonen, J., Repo, M., and Laakasuo, M. (2018). Evaluating the replicability of the uncanny valley effect. *Heliyon*, 4(11):e00939.

- 
- Phillips, E., Zhao, X., Ullman, D., and Malle, B. F. (2018). What is Human-like? Decomposing Robots' Human-like Appearance Using the Anthropomorphic roBOT (ABOT) Database. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 105–113.
- Piçarra, N. J. G. (2014). *Predicting intention to work with social robots*. PhD thesis, Universidade do Algarve.
- Piwek, L., McKay, L. S., and Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, 130(3):271–277.
- Pochwatko, G., Giger, J.-C., Róžańska-Walczuk, M., Świdrak, J., Kukielka, K., Możaryn, J., and Piçarra, N. (2015). Polish version of the negative attitude toward robots scale (NARS-PL). *Journal of Automation Mobile Robotics and Intelligent Systems*, 9:65–72.
- Pollick, F. E. (2009). In search of the uncanny valley. In *International Conference on User Centric Media*, pages 69–78, Venice, Italy.
- Ratajczyk, D. (2019). Uncanny valley in video games: An overview. *Homo Ludens*, 12(1):135–148.
- Ratajczyk, D., Jukiewicz, M., and Łupkowski, P. (2019). Evaluation of the uncanny valley hypothesis based on declared emotional response and psychophysiological reaction. *Bio-Algorithms and Med-Systems*, 15(2).
- Robinson, H., MacDonald, B., and Broadbent, E. (2014). The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics*, 6(4):575–591.
- Rotshtein, P., Malach, R., Hadar, U., Graif, M., and Hendler, T. (2001). Feeling or features: Different sensitivity to emotion in high-order visual cortex and amygdala. *Neuron*, 32(4):747–757.
- Rozin, P. and Fallon, A. E. (1987). A perspective on disgust. *Psychological review*, 94(1):23.
- Schein, C. and Gray, K. (2015). The eyes are the window to the uncanny valley: Mind perception, autism and missing souls. *Interaction Studies*, 16(2):173–179.
- Schindler, S., Zell, E., Botsch, M., and Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports*, 7:45003.

- Seyama, J. and 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.
- Sharkey, A. and Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics and information technology*, 14(1):27–40.
- Stein, J.-P., Liebold, B., and Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior*, 95:73–82.
- Stein, J.-P. and Ohler, P. (2017). Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160:43–50.
- Strojny, P. and Strojny, A. (2014). Kwestionariusz immersji—polska adaptacja i empiryczna weryfikacja narzędzia. *Homo Ludens*, 1(6):171–186.
- Tay, T., Low, R., Loke, H., Chua, Y. L., and Goh, Y. (2018). Uncanny valley: A preliminary study on the acceptance of Malaysian urban and rural population toward different types of robotic faces. In *IOP Conference Series: Materials Science and Engineering*, volume 344, page 012012. IOP Publishing.
- Thellman, S., Silvervarg, A., and Ziemke, T. (2017). Folk-psychological interpretation of human vs. humanoid robot behavior: Exploring the intentional stance toward robots. *Frontiers in psychology*, 8:1962.
- Thepsonthorn, C., Ogawa, K.-i., and Miyake, Y. (2021). The exploration of the uncanny valley from the viewpoint of the robot’s nonverbal behaviour. *International Journal of Social Robotics*, 13:1443–1455.
- Thompson, P. (1980). Margaret thatcher: A new illusion. *Perception*, 9(4):483–484.
- Thornhill, R. and Gangestad, S. W. (1999). Facial attractiveness. *Trends in cognitive sciences*, 3(12):452–460.
- Tinwell, A. (2014). *The uncanny valley in games and animation*. AK Peters/CRC Press.
- Tinwell, A., Grimshaw, M., Nabi, D. A., and Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior*, 27(2):741–749.

- 
- Wang, S., Lilienfeld, S. O., and Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4):393–407.
- Yam, K. C., Bigman, Y., and Gray, K. (2021). Reducing the uncanny valley by dehumanizing humanoid robots. *Computers in Human Behavior*, 125:106945.
- Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research*, 55(1):20–32.
- Złotowski, J., Strasser, E., and Bartneck, C. (2014). Dimensions of anthropomorphism: from humanness to humanlikeness. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 66–73.
- Złotowski, J., Yogeeswaran, K., and Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, 100:48–54.
- Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., and Ishiguro, H. (2015). Persistence of the uncanny valley: The influence of repeated interactions and a robot’s attitude on its perception. *Frontiers in psychology*, 6:883.

## Appendix A

# Appendices

### A.1 Study I: Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments



# Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments

Dawid Ratajczyk<sup>1</sup>

Accepted: 27 June 2022 / Published online: 16 August 2022  
© The Author(s) 2022

## Abstract

The uncanny valley hypothesis (UVH) suggests that almost, but not fully, humanlike artificial characters elicit a feeling of eeriness or discomfort in observers. This study used Natural Language Processing of YouTube comments to provide ecologically-valid, non-laboratory results about people’s emotional reactions toward robots. It contains analyses of 224,544 comments from 1515 videos showing robots from a wide humanlikeness spectrum. The humanlikeness scores were acquired from the Anthropomorphic roBOT database. The analysis showed that people use words related to eeriness to describe very humanlike robots. Humanlikeness was linearly related to both general sentiment and perceptions of eeriness—more humanlike robots elicit more negative emotions. One of the subscales of humanlikeness, Facial Features, showed a UVH-like relationship with both sentiment and eeriness. The exploratory analysis demonstrated that the most suitable words for measuring the self-reported uncanny valley effect are: ‘scary’ and ‘creepy’. In contrast to theoretical expectations, the results showed that humanlikeness was not related to either pleasantness or attractiveness. Finally, it was also found that the size of robots influences sentiment toward the robots. According to the analysis, the reason behind this is the perception of smaller robots as more playable (as toys), although the prediction that bigger robots would be perceived as more threatening was not supported.

**Keywords** Uncanny valley · Robot · YouTube · Emotions · Text mining · Social robotics · Natural Language Processing

## 1 Introduction

The uncanny valley hypothesis (UVH) suggests that almost, but not fully, humanlike artificial characters will elicit a feeling of eeriness or discomfort in observers (see [35,47]). This characteristic drop in likability is called the uncanny valley. Such an effect is considered within the field of humanoid robotics and also for computer-generated imagery (e.g., [20]). The concept of the uncanny valley has gained much attention in recent years; however, there are still certain inconsistencies in the debate. The doubts address not only the explanations for and the depth of the uncanny valley but also the dependent variable, which is commonly (but not exclusively) referred to as the affinity dimension [25]. These issues are strongly related to ambiguities with the emo-

tions related to the uncanny valley [25,62]. Kätsyri et al. [25] suggest that terms used in the uncanny valley studies (eeriness, likability, familiarity, and affinity) are related to various aspects of perceptual familiarity and emotional valence, and that the “empirical studies would be necessary for resolving which self-report items would be ideal for measuring affinity” (p. 3). There have been a few attempts to disambiguate self-report language describing the emotions in the uncanny valley within experimental studies (e.g., [20,21]); however, the results are not conclusive, and the uncanny valley research constantly does not use consistent language. The dependent variable has variously been considered as, for example, perceived warmth [27], eeriness, pleasantness, and creepiness [26], acceptance [54], or comfort level [30,43]. For a wider discussion, see Wang et al. [62, pp. 398–399].

Another point is that explanations of the uncanny valley focus robustly on the visual aspects of robots (e.g., [7,8,31,41,47]). However, recent research shows that negative or positive emotions toward identical artificial agents can be moderated by the individual’s belief as to whether the agents are directed by artificial or human intelligence [50].

---

Dawid Ratajczyk  
dawid.ratajczyk@amu.edu.pl

<sup>1</sup> Faculty of Psychology and Cognitive Science, Adam Mickiewicz University in Poznan, Szamarzewskiego 89/AB, 60–568 Poznań, Poland

Given the lack of agreement on what causes the uncanny effect, the involvement of variables additional to visual aspects (such as movement, behavioral social cues, proximity of agent, and others) seems plausible.

Most studies regarding UVH try to elicit the emotions in experimental conditions and measure them using ad hoc questionnaires (e.g., [39]) or unspecific, general-feeling questionnaires (e.g., [10]). Such an approach, despite the obvious advantages of controlled experiments, has negative implications for studying emotions. Aside from the explicit influence of an experiment on participants' emotions, the mood in which subjects walk into the laboratory has a large effect on positive and negative affect induction [14] and this may potentially lead to distorted results.

The sinusoidal UVH relationship introduced by Mori [35] is not unambiguously reflected in empirical research. The name of the UVH phenomenon (i.e., the “valley”) refers to the specific graph shape of the emotional reaction toward humanlike agents – sharp decrease and increase of affinity. However the empirical evidence for such a curve is sparse. An extensive review by Kätsyri et al. [25] demonstrated a linear relationship between affinity and the humanlikeness, i.e., affinity reaction increased proportionally to the humanlikeness. A few other graphs have been considered, i.e., U-shaped relationship (e.g., [29,45]), and cliff-like relationship (e.g., [3]). Since the UVH was defined by the shape of the relationship between humanlikeness and affinity, there is a strong need to resolve this issue.

Considering the above, the reasons for lack of agreement in UVH research may be divided into: (a) inconsistencies in affinity dimension assessment, (b) limitation of research stimuli to visual aspects; and (c) difficulties in emotion elicitation in laboratory conditions. To address these issues I tested emotions and language in more ecological, non-laboratory conditions using Natural Language Processing (NLP) of comments on robots videos in social media. The novelty of the approach relies on studying more natural human reactions than those in surveys and experiments, as well as objectivization of the emotion assessment method. An analysis of large samples of natural human expressions using automated text processing allows to evaluate how various dependent variables (i.e., eeriness, pleasantness, and attractiveness) are associated with humanlikeness, possibly resulting in unraveling of some inconsistencies (addressing issue (a)). The present study exploits various videos of robots, hence it takes into account variables that are absent in 2D stimuli presentation, such as behavior of robots, their voice, size and others (addressing issue (b)). In contrast to experimental conditions, people using the Internet as part of their everyday life may manifest more natural emotional reactions toward the robots that they encounter. Therefore, an analysis of such manifestations allows to research genuine UVH-related reactions (addressing issue (c)).

During the last few years, social media has grown to become highly popular, providing not only a means of communication between people but also seizing control of many more social activities, such as the creation of reputation or the enabling a social life [36]. In order to study attitudes toward robots in popular media, I used comments on robot videos from the YouTube video-sharing platform. YouTube is a highly popular internet service (ranked #2 in global internet engagement according to *alexa.com*<sup>1</sup>). Previous studies show that the analysis of YouTube comments allows community opinions and also emotions toward a specific topic to be determined [18,49]. One of the methods used to study affective states in people's statements is sentiment analysis. The method allows information to be gathered about attitudes, emotions, and opinions, and it is widely used in social media data extraction [1,5].

The structure of this paper is as follows. In the further part of the Introduction, I present studies that are relevant to the research objectives and formulate research hypotheses. In the Method section, I describe acquisition and processing of comments, and also selection of emotional and humanlikeness indicators. In the Results section, I test the hypotheses and perform additional exploratory analysis. In the Discussion, I interpret the results, generalize the findings, and consider the limitations of the study. The paper is supplemented with an Appendix, where the list of all analyzed robots and their scores is presented.

## 1.1 Related Work

Among the work with the closest methodological approach to this research, several studies have examined online comments related to the evaluation of robots. One of the first ones in this field was carried out by Friedman et al. [15]. They investigated online discussion forums associated with robotic dog AIBO. During the study, 3119 postings were coded with subsequent general categories: technological essences, life-like essences, mental states, social rapport, and moral standing. Results showed that AIBO psychologically engaged its owners, and people created relationships with the robotic pet. The robot evoked conceptions of life-like essences, mental states, and social rapport, and sporadically conceptions of moral standing. The authors considered that playfulness might have a part in language people use to refer to their robotic pets, as it influences users actual emotions and thoughts.

Strait et al. [52] performed an analysis of YouTube comments regarding robots using the raters method, which involved an evaluation of comments' topics by assistants. They investigated public perception toward groups of

<sup>1</sup> <https://www.alexa.com/siteinfo/youtube.com>

[Accessed: 07.12.2020].

mechanomorphic and humanlike robots and established a less positive valence and more UVH-related comments toward humanlike robots compared with mechanomorphic ones. They also discovered that decreased emotional valence in comments is partially related to fears of a “technology takeover” (i.e., a fear of robots replacing humans). Additionally, they stressed the occurrence of the widespread sexualization and objectification of the female-gendered robots in comments. However, this study has certain limitations. Their analysis involved comments from subjectively chosen videos and also the subjective exclusion of some comments. The number of comments analyzed by judges was rather small (1200) and their methodology did not enable them to perform a systematic inquiry of emotional words used for the description of uncanny characters.

Hover et al. [22] extended the work of Strait et al. [52] and used a similar methodological approach to the analysis of comments regarding more and less humanlike robots. Their results also indicated that more humanlike robots elicit more negative emotional reactions related to the uncanny valley effect. They also confirmed the results that female humanlike robots were more likely to be subject to sexualization and sexism than male robots. Their data suggests that there is more difference in examined factors between less and more humanlike robots than female and male robots. Also, less humanlike robots were more likely to evoke perceptions of a threat than highly humanlike robots.

Strait et al. [51] tested racialization toward robots with appearances of different racial identities based on analyses of YouTube comments. They used comments from videos of a Black, White, and Asian appearing robots and humans (6 videos in total). The results show that people extend and amplify racial biases toward robots, and also that dehumanization based on social stereotypes is greater for robots.

Vlachos and Tan [59], instead of using human annotators, did an analysis of YouTube comments regarding four humanlike robots (involving comments on four videos), with the utilization of text mining and machine learning. Their work was entirely exploratory, not focused explicitly on the uncanny valley, but rather on a general interaction with highly humanlike robots. The authors distinguished three topics important for robotics: human–robot relationships, technical specifications, and the so-called science fiction valley (a combination of the UVH concept and references to science fiction movies and games). The limitations of this study were the choice of only four videos, a lack of manipulation of humanlikeness, and the involvement of replies to main comments, which may contain off-topics.

Also, Yu [63] studied attitudes toward robots employed as hotel workers. They collected comments on two YouTube videos and coded them automatically in reference to concepts related to the perception of robots (anthropomorphism, animacy, likability, perceived intelligence, and perceived

safety). Their cluster analyses of data showed that likability and anthropomorphism are the most distinct concepts. The results supported the existence of the uncanny valley. The discomfort in form of anxiety feeling co-occurred with perceived intelligence. Additionally, discussions about movement of robots were related to machinelikeness. Although the thematic analysis was automatic, the cleaning of data was done manually. Therefore, the sample size of comments and videos was rather small and videos were limited to robots from a very specific context.

Considering the variables, that may have influence on robot perception, the results of Wang et al. [61] show that the size of agents which are otherwise visually equivalent determines the degree to which they are perceived as uncanny. Participants in augmented reality preferred smaller virtual agents over visually identical human-size agents, referring to these as too large, imposing, weird, and creepy. These findings are in line with the conclusions of the analysis of Kätsyri et al. [25] pointing out that the uncanny valley concept is in fact very complex and suggest there is a need for a closer examination of the influence of robot size on UVH-related feelings. Also, Mori [35] pointed that higher humanlikeness may be perceived when absolute size of agent is ignored. Wang et al. [61] suggested, on the bases of subjects’ feedback, that small embodied agents are more entertaining and amusing than other agents. Also, Wagner et al. [60] reported that fun plays an important role in embodied agents. As such, people may treat smaller robots as if they were playable or related to fun. Another plausible explanation is that bigger robots can be seen as stronger and more threatening to people, and therefore evoke negative emotions toward them.

In the following study, I use methods of data acquisition and processing that are automated and resilient to the possibility of a biased, subjective choice of movies and comments. I acquire a large number of utterances referring to robots with different humanlikeness and conduct several analyses in order to exploit the potential of the data.

The aims of this paper are as follows: (1) to test the relationship between robot humanlikeness and sentiment scores, (2) to examine which of the variables (eeriness, pleasantness, and attractiveness) are related to humanlikeness with the new NLP method, (3) to test the impact of robot size on sentiment and to examine the reasons behind the observed relationship, and (4) to characterize the specific, emotional words expressed toward robots.

Additionally, I investigated the awareness of UVH among commenters. The popularity of the UVH concept among internet communities seems to be widespread, as indicated in popular articles and by an extensive list of spotted uncanny

valley examples in animations and video games prepared by users<sup>2</sup>.

## 1.2 Hypotheses

On the basis of the above-mentioned literature, the following hypotheses have been formulated:

**H1** *The shape of the graph representing the relationship between humanlikeness (and its subscales) and sentiment valence toward robots is linear.*

**H2** *Emotional indicators (eeriness, pleasantness, and attractiveness) are equally related to humanlikeness. The relationships between humanlikeness (and its subscales) and emotional indicators are linear.*

**H3** *The size (i.e., height) of robots has an impact on emotions elicited by robots.*

**H3a** *The smaller a robot, the more it is perceived as playable or related to fun.*

**H3b** *The bigger a robot, the more it is perceived as threatening and dangerous.*

## 2 Methods

The work included: (1) data retrieval (downloading the comments regarding robots from the YouTube platform), (2) processing of comments (cleaning the text and extracting emotional indicators), and (3) acquiring humanlikeness scores for robots.

### 2.1 YouTube Comments Collection

The method of data collection was inspired by the publication of Thelwall [55]. The method allows videos relevant to a given topic to be systematically searched for and acquired without engaging subjective preferences.

The topic of the investigation was existing robots. In order to acquire utterances regarding robots from a humanlikeness spectrum, I prepared a list of 246 developed and functional robots—242 robots from the Anthropomorphic roBOT (ABOT) Database<sup>3</sup> plus 4 additional robots which were not included in the ABOT database<sup>4</sup>.

<sup>2</sup> <https://www.wired.com/story/uncanny-valley-robot-voices/>; <https://vtropes.org/pmwiki/pmwiki.php/Main/UncannyValley> [Accessed: 07.12.2020].

<sup>3</sup> <http://abotdatabase.info/collection> [Accessed: 31.08.2020]; 9 robots were excluded from the original 251 item list due to nonspecific names and potential collisions during YouTube search.

<sup>4</sup> Robots added: Vector by Anki, Spot by Boston Dynamics, earlier and later version (after 28.03.2019) of Handle by Boston Dynamics. They were added in order to maximize the number of robots for analysis.

I adopted the API protocol shared by YouTube and wrote a Python 3.8 script in order to download comments related to a particular robot. Firstly, I acquired the relevant videos list for each robot. The criteria of the search for videos were as follows. I included only short videos (less than 4 min<sup>5</sup>) in order to focus on robots' presentations and reduce the possibility of the occurrence of uncontrolled variables such as the presentation of multiple robots or excessive commentary in the video, for example. The relevance language (the API option) was English and the region of the search was the US<sup>6</sup>. Videos (and comments) were not limited by date. The search phrase was combined from the robot's name, the word 'robot' and the additional clue (the name of the production company or creator, country where the robot was developed, or the word 'humanoid'). The phrases were prepared in order to maximize the number of relevant videos. All the search phrases are included in the Supplementary Materials<sup>7</sup>. The comments scraping procedure was performed between 1st and 10th August 2020.

After selecting videos with the described method, I automatically evaluated their accuracy according to the following criterion: I left only the videos (2157 in total from original 8782) which had the name of the given robot in the title<sup>8</sup>. After this, I downloaded all of the comments for the listed videos. Concerning the relevance to the topic of the videos (robots), I discarded replies to comments and left only primary comments. I removed empty comments and those duplicated with more than 100 characters (long duplicated comments might be spam or created by bots). Then I removed non-English comments, which were detected with the use of the Python *langdetect* v1.0.7 package (<https://pypi.org/project/langdetect/>). The total number of comments after processing was 228,688 from 2149 videos. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

### 2.2 Analysis of Comments

Each comment was processed with Python 3.8 script<sup>9</sup>. Firstly, all of the hyperlinks were removed from the comments and the comments were part-of-speech tagged with the use of the *NLTK POS tagger*. Then the stopwords, punctuation,

<sup>5</sup> The YouTube API allows videos below 4, between 4 and 20, and above 20 minutes to be searched for.

<sup>6</sup> Due to the YouTube API features, this does not exclude all the other videos (with different languages or from different regions).

<sup>7</sup> The Supplementary Materials are available at <https://osf.io/cyvmx/>.

<sup>8</sup> The reason behind this is that the YouTube search algorithm is not straightforward and may decide that a video is relevant even if there is no keyword in the title.

<sup>9</sup> The following packages were used: *string*, *NLTK*, *AFINN*, *collections*, *NumPy*, and *pandas*.

**Table 1** Example of counting the sentiment scores for comments with *AFINN* package

	Original comment	Processed comment with word scores	Score
1.	“This is awesome. Amazing!”	Awesome [4], amazing [4]	8
2.	“It’ll be scary when they put guns in their hands”	Scary [− 2], put [0], gun [− 1], hand[0]	− 3
3.	“It’s ridiculous, rough, ugly and so more adjectives.”	Ridiculous [− 3], rough [0], ugly [− 3], adjective [0]	− 6
4.	“Does that robot vaccum too?”	Robot [0], vaccum [0]	0

and non-alphabetic words were removed. All the remaining words were lemmatized.

In order to obtain reliable data for further analysis, I used only those robots that have more than 200 suitable (not discarded by the above criteria) comments in total. The same cut-off number was used by Guzman et al. [17] in a similar sentiment analysis of comments from the GitHub platform. The selected number is a trade-off between an insufficient number of comments for unbiased analysis and a sufficient number of robots for further analysis (see elaboration in the Limitations section). That left 33 robots suitable for further analysis (224,544 comments from 1515 videos in total). List of these robots is presented in the “Appendix”.

*Sentiment score* The sentiment scores were counted with the use of the *AFINN* v0.1 Python package (<https://pypi.org/project/afinn/>), which provides a lexicon of emotional words with scores ranging from − 5 to 5. The lexicon was prepared on the basis of internet fora and microblogs, and contains internet slang and obscene words [37], making it suitable for a YouTube comments analysis. The score for each comment was counted by adding up individual scores for every word in the processed comment (see example in Table 1). The mean from all comments referring to a robot was then counted in order to obtain the robot sentiment score. All the scores for individual robots are presented in the “Appendix”.

*Eeriness, pleasantness, and attractiveness indices* Additionally, I have defined indices to distinguish emotional terms characteristic to the uncanny valley, i.e., related to emotions elicited by observation or contact with close to human-like agents. The reason for choosing concepts of eeriness, pleasantness, and attractiveness is that these are the most discussed dependencies in the context of UVH studies (e.g., [20,25,57]). There are tools available for accurate emotions identification in language such as Linguistic Inquiry and Word Count (LIWC) [53]. However, the area of the uncanny valley is very specific and, although there were attempts to disambiguate the words used for naming emotions toward robots at least for eeriness (e.g., [21,48]), in order to maintain consistency among examined concepts, I created original word sets for identifying UVH-related emotions. I created three lists of words related to each aforementioned concept

**Table 2** All the words counted for eeriness, pleasantness, attractiveness, and familiarity indices

Index	Words related to the index
Eeriness	Creepy, eerie, haunting, spookish, spooky, uncanny, unearthly, weird
Pleasantness	Agreeable, blessed, congenial, darling, delectable, delicious, delightful, delightsome, dreamy, dulcet, enjoyable, felicitous, good, grateful, gratifying, heavenly, jolly, luscious, nice, palatable, pleasant, pleasing, pleasurable, pretty, satisfying, savory, sweet, tasty, welcome
Attractiveness	Aesthetic, attractive, beauteous, beautiful, bonny, comely, cute, drop-dead, fair, fetching, good, good-looking, goodly, gorgeous, handsome, knockout, likely, lovely, lovesome, pretty, ravishing, seemly, slightly, stunning, taking, well-favored

using the Merriam-Webster Online Thesaurus<sup>10</sup>. As the list of videos acquired from YouTube was targeted for the US, I used the American dictionary, and this one was used previously by Kätsyri et al. [25] for defining the concepts related to the uncanny valley. The following definitions for each concept were used when identifying synonyms: eerie—fearfully and mysteriously strange or fantastic; pleasant—giving pleasure or contentment to the mind or senses; attractive—very pleasing to look at. Lists of all synonyms for each index are presented in Table 2. Afterward, eeriness, pleasantness, and attractiveness indices were counted for each robot. Establishing an index for a given robot required the following steps. I counted the relative frequencies of words related to each concept. I excluded the counts of the word ‘uncanny’ which occurred in the phrase ‘uncanny valley’ in order to focus on expressions of emotions, not the awareness of the phenomenon. This method gave me the possibility of a systematic, numerical evaluation of each concept for every robot. All the scores for individual robots are presented in the “Appendix”.

<sup>10</sup> <https://www.merriam-webster.com/thesaurus>  
07.12.2020].

[Accessed:

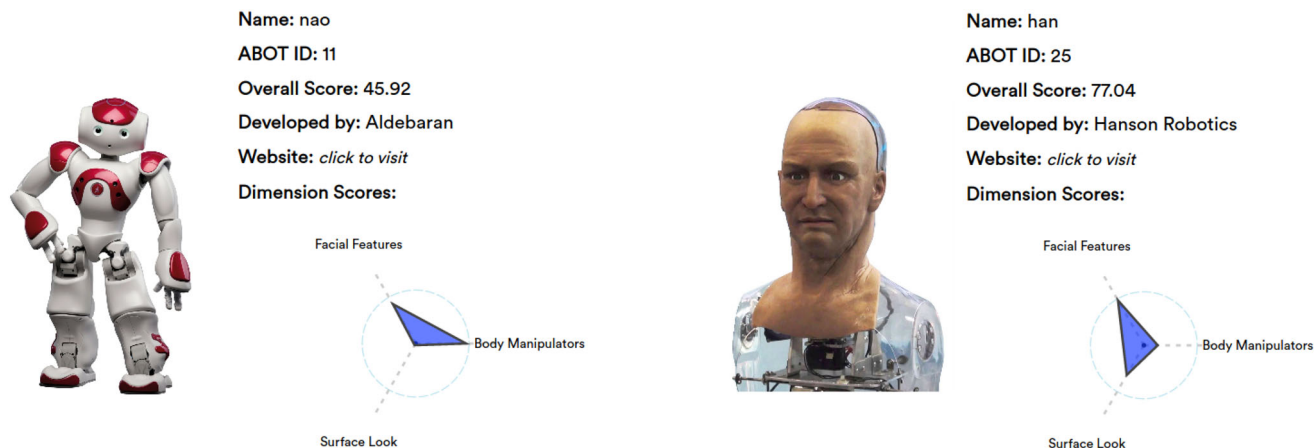


Fig. 1 Illustrative robots from ABOT database with humanlikeness (and subscales) scores

## 2.3 Humanlikeness of Robots

Appraisals of robots' humanlikeness were acquired from the ABOT database, and using a tool provided by the database creators [40]. The ABOT database is an open source collection of real-world robots with their humanlikeness scores. The database allows the unification of robots' humanlikeness dimension among studies and the investigation of the impact of distinguished underlying factors.

Phillips et al. [40] created the ABOT database with humanlikeness scores and related factors based on a study using a collection of 200 images of real-world robots. They uncovered three distinct appearance dimensions (i.e., bundles of features) that contribute to the anthropomorphism of robots. They distinguished the following subscales: (1) Surface Look (presence of eyelashes, head hair, skin, genderedness, nose, eyebrows, apparel), (2) Body-Manipulators (presence of hands, arms, torso, fingers, legs), and (3) Facial Features (presence of face, eyes, head, mouth). The three subscales positively correlate with humanlikeness. Exemplary robots from ABOT database are presented in Figure 1.

The ABOT authors shared scores of general humanlikeness and its subscales for 251 robots and also made available a tool for assessing the humanlikeness of robots not present in the database. Because the ABOT scales were created on a large sample of participants (over 1000), it seems to be an appropriate appliance for the unification of robots' humanlikeness dimension among studies. These scores will be used for further analyses. All the ABOT scores for robots used for this study are presented in the Appendix.

In what follows, I will use the main ABOT humanlikeness score but also all of its subscales.

## 3 Results

### 3.1 Humanlikeness and Sentiment Scores

The UVH describes a non-linear relationship between humanlikeness and emotional reaction. Firstly, I tested the relationship between humanlikeness of robots and general sentiment scores and examined how the relationship changes for particular ABOT subscales of humanlikeness (H1). Therefore, in order to characterize the shape of the relationship and test which model better describes the relationship (linear, quadratic or cubic), I performed polynomial curve fitting and tested the goodness of fit of each regression model using the Akaike Information Criterion (AIC; see [9]). AIC allows us to compare models of varying complexity, penalizing a higher number of parameters. For small sample sizes ( $n/K < 40$ , where  $n$  is sample size and  $K$  is number of parameters), as in this case, Burnham and Anderson [9, p. 66]) suggested the usage of the adjusted formula:

$$AIC_c = n * \ln \left( \frac{RSS}{n} \right) + 2 * K + \frac{2 * K * (K + 1)}{n - K - 1},$$

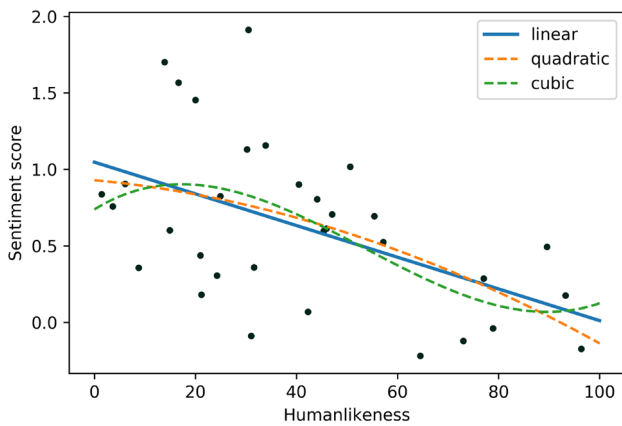
where  $RSS$  is residual sum of squares. A model with the lowest  $AIC_c$  is preferred. Also,  $R^2$  was counted in order to examine the variance in the dependent variable that is predicted from the independent variable. The results are presented in Table 3.  $P$  values were corrected with the Benjamini–Hochberg adjustment (see [6]) for multiple comparisons (12 tests).

The results show that, out of the three models, the linear one was the one which best fit the relationship between humanlikeness and sentiment. As for the ABOT subscales of humanlikeness, Surface Look also has a linear relationship with sentiment. However, the relationship between Facial Features and sentiment was best represented by the cubic model. Sentiment was moderate at very low humanlikeness,

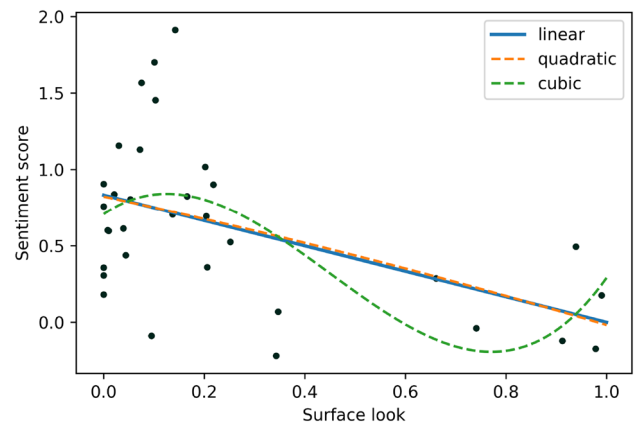
**Table 3** Polynomial regression comparison between models of the humanlikeness (and subscales) and sentiment relationship

Dependent var.	Independent var.	Model	$AIC_c$	$\Delta_i(AIC)$	$R^2$	$F$	$p$
Sentiment	Humanlikeness	<b>Linear</b>	<b>-47.02</b>	<b>0.0</b>	<b>0.26</b>	<b>10.81</b>	<b>0.015</b>
		Quadratic	-45.05	2.0	0.27	5.52	0.022
		Cubic	-43.38	3.6	0.29	3.93	0.027
	Surface look	<b>Linear</b>	<b>-46.2</b>	<b>0.0</b>	<b>0.24</b>	<b>9.78</b>	<b>0.015</b>
		Quadratic	-43.79	2.4	0.24	4.74	0.027
		Cubic	-44.48	1.7	0.31	4.39	0.023
	Facial features	Linear	-37.25	22.3	0.0	0.1	0.757
		Quadratic	-46.15	13.4	0.29	6.21	0.017
		<b>Cubic</b>	<b>-59.52</b>	<b>0.0</b>	<b>0.56</b>	<b>12.51</b>	<b>&lt;0.001</b>
Body-Manipul.	Linear	-38.01	0.0	0.03	0.82	0.497	
	Quadratic	-36.1	1.9	0.04	0.64	0.641	
	Cubic	-33.52	4.5	0.04	0.42	0.757	

Bold, underlined rows indicate best, significant models for particular relationships



**Fig. 2** Sentiment score and Humanlikeness scale relationship curve fitting

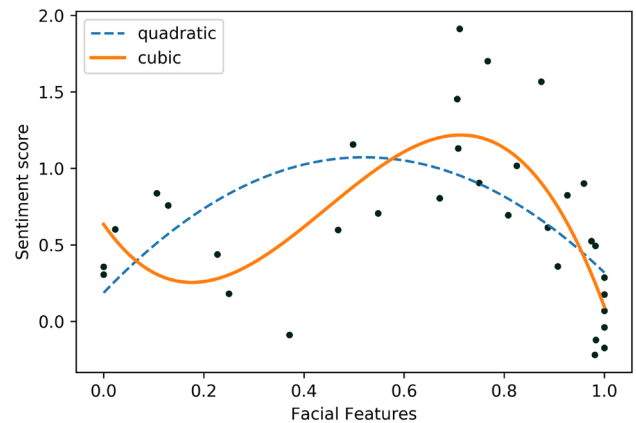


**Fig. 3** Sentiment score and Surface Look subscale relationship curve fitting

decreased at low humanlikeness, was highest at high humanlikeness, but lowest at very high humanlikeness. For the Body-Manipulators subscale, none of the models was significant. Plots for significant models are presented in Figs. 2, 3, and 4. The best models were drawn with a solid line. The results partially support H1—out of 4 examined relationships, 2 were linear. Generally, the more humanlike robots are (and the more humanlike surface they have), the more negative sentiment they elicit.

**3.2 Humanlikeness and Emotional Indicators**

In order to test the associations between particular emotional indicators (eeriness, pleasantness, and attractiveness) and humanlikeness (H2) I conducted a similar polynomial regression analysis as for the relationship between humanlikeness and sentiment. For each indicator the  $AIC_c$  values,  $R^2$ , and significance values for models (with Benjamini–Hochberg



**Fig. 4** Sentiment score and Facial Features subscale relationship curve fitting

adjustment for 12 tests) were calculated. Results are shown in Table 4.

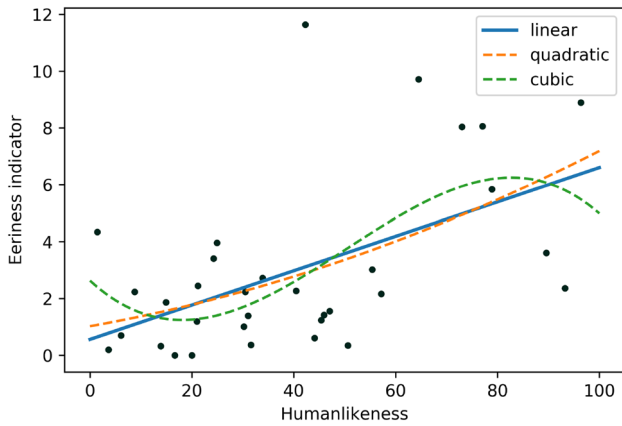
**Table 4** Polynomial regression comparison between models of the humanlikeness (and subscales) and emotional indicators relationships

Dependent var.	Independent var.	Model	$AIC_c$	$\Delta_i(AIC)$	$R^2$	$F$	$p$	
Eeriness	Humanlikeness	<b>Linear</b>	<b>66.12</b>	<b>0.0</b>	<b>0.28</b>	<b>11.94</b>	<b>0.003</b>	
		Quadratic	68.32	2.2	0.28	5.92	0.012	
		Cubic	68.78	2.66	0.33	4.72	0.013	
	Surface Look	Linear	63.25	3.88	0.34	15.84	0.001	
		Quadratic	60.96	1.59	0.43	11.15	0.001	
		<b>Cubic</b>	<b>59.37</b>	<b>0.0</b>	<b>0.49</b>	<b>9.47</b>	<b>0.001</b>	
	Facial Features	Linear	72.31	14.92	0.13	4.59	0.053	
		Quadratic	62.06	4.67	0.41	10.29	0.001	
		<b>Cubic</b>	<b>57.39</b>	<b>0.0</b>	<b>0.52</b>	<b>10.65</b>	<b>0.001</b>	
	Body-Manipul.	Linear	76.76	0.0	0.0	0.11	0.839	
		Quadratic	78.72	1.97	0.02	0.26	0.839	
		Cubic	81.32	4.56	0.02	0.17	0.914	
	Pleasantness	Humanlikeness	Linear	35.76	0.04	0.14	5.13	0.146
			Quadratic	35.72	0.0	0.2	3.84	0.146
			Cubic	36.31	0.59	0.25	3.24	0.146
Surface Look		Linear	37.86	0.0	0.09	2.91	0.295	
		Quadratic	39.11	1.25	0.12	2.0	0.298	
		Cubic	40.28	2.42	0.16	1.78	0.298	
Facial Features		Linear	40.62	0.0	0.01	0.18	0.671	
		Quadratic	41.1	0.48	0.06	1.01	0.412	
		Cubic	42.02	1.4	0.11	1.19	0.412	
Body-Manipul.		Linear	39.87	0.56	0.03	0.9	0.412	
		Quadratic	39.31	0.0	0.11	1.9	0.298	
		Cubic	41.91	2.6	0.11	1.23	0.412	
Attractiveness		Humanlikeness	Linear	65.99	0.0	0.03	1.1	0.697
			Quadratic	67.9	1.91	0.05	0.78	0.697
			Cubic	69.19	3.2	0.09	0.91	0.697
	Surface Look	Linear	67.13	0.0	0.0	0.01	0.929	
		Quadratic	69.41	2.28	0.0	0.07	0.929	
		Cubic	69.38	2.25	0.08	0.85	0.697	
	Facial Features	Linear	66.7	0.0	0.01	0.42	0.697	
		Quadratic	68.75	2.05	0.02	0.38	0.825	
		Cubic	69.54	2.84	0.08	0.8	0.697	
	Body-Manipul.	Linear	63.37	0.0	0.11	3.76	0.697	
		Quadratic	65.78	2.41	0.11	1.83	0.697	
		Cubic	68.2	4.83	0.11	1.24	0.697	

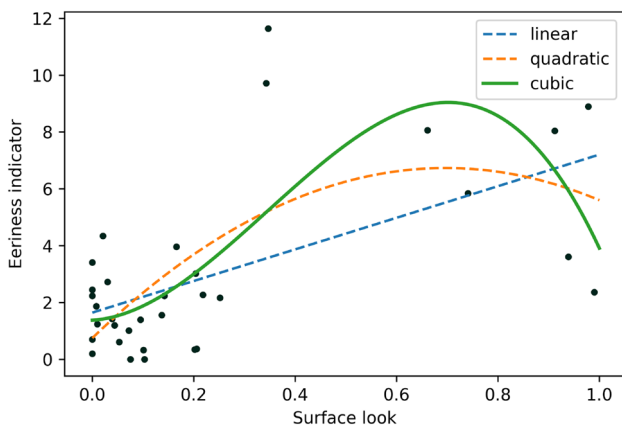
Bold underlined rows indicate best, significant models for particular relationships

The results indicate that for the general humanlikeness and eeriness relationship, the best model is linear. However, for the Surface Look and Facial Features subscales, the best models are cubic. As for pleasantness and attractiveness, none of the models was significant. Plots of significant models are presented in Figs. 5, 6, and 7. Therefore, the first part of H2 should be rejected—emotional indicators are not equally

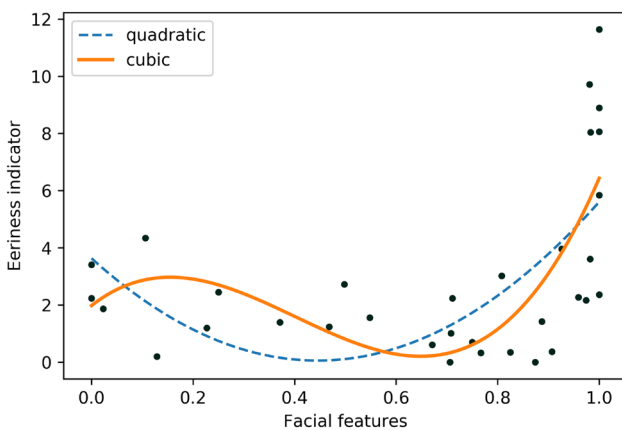
related to the humanlikeness. Eeriness is the most important factor for humanlikeness. The results partially support the second part of H2—the relationship between humanlikeness and the only significant index (eeriness) is linear. However, none of humanlikeness subscales affect the eeriness linearly.



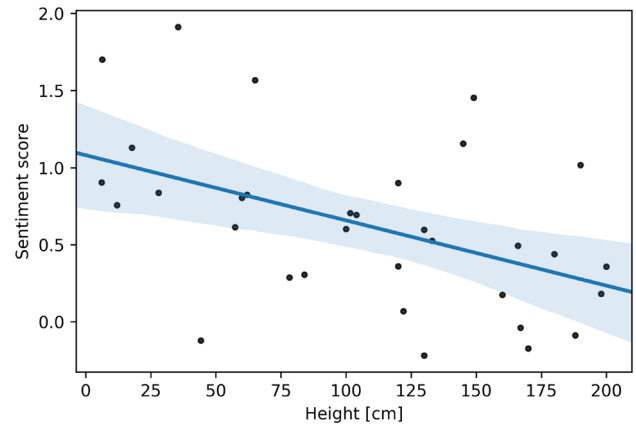
**Fig. 5** Eeriness index and Humanlikeness scale relationship curve fitting



**Fig. 6** Eeriness index and Surface Look subscale relationship curve fitting



**Fig. 7** Eeriness index and Facial Features subscale relationship curve fitting



**Fig. 8** Regression plot of height of robots and sentiment score relationship

### 3.3 Height of the Robot and Sentiment Score

Literature regarding UVH subject suggests that the size of robots may have an impact on emotions elicitation [35,61]. It may be a confounding factor in the human–robot interaction (HRI) studies. Nevertheless, it seems that the relationship has not been tested empirically. Therefore, I tested the influence of robots’ height on general sentiment score (H3). The size of robots was acquired from documentations shared by producers, promotional materials, and articles about robots. For two robots (Han, Bina48), the information about their size was not available, therefore I used multiple independent raters method to evaluate their height. 6 experts in social robotics scrutinized pictures and movies of the robots, and estimated the height on the basis of the comparison to various elements visible in the scenes (humans, computers, other elements that their found useful). The mean of their responses was taken as the height<sup>11</sup>.

The regression model for predicting the sentiment by robot size was significant ( $F(31, 1) = 9, p = 0.005$ ),  $R^2$  was equal to 0.23. The size coefficient was equal to  $\beta = -0.004$  ( $t(31) = -3, p = 0.005$ ). The regression plot is shown in Figure 8. This result supports the H3, indicating that the height of a robot has a significant impact on general emotions elicited by robots. The bigger a robot is, the more negative emotions it elicits.

### 3.4 Explanation of Sentiment Score and Robots’ Height Relation

In order to find out what is the reason behind the observed relationship between sentiment score and robots’ height, hypotheses H3A and H3B have been formulated. H3A states

<sup>11</sup> Estimated height for Han: 78.3 cm ( $SD = 14.6$ ); for Bina48: 44.2 cm ( $SD = 4.5$ ).

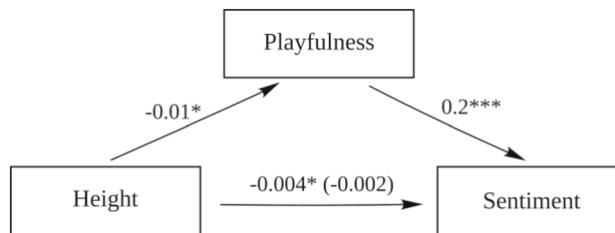
**Table 5** All the words counted for the playfulness and threateningness indices

Index	Words related to the index
Playfulness	Dalliance, frolic, frolicking, fun, play, recreation, relaxation, rollicking, sport
Threateningness	Dangerous, grave, grievous, hazardous, jeopardizing, menacing, parlous, perilous, risky, serious, threatening, unhealthy, unsafe, venturesome

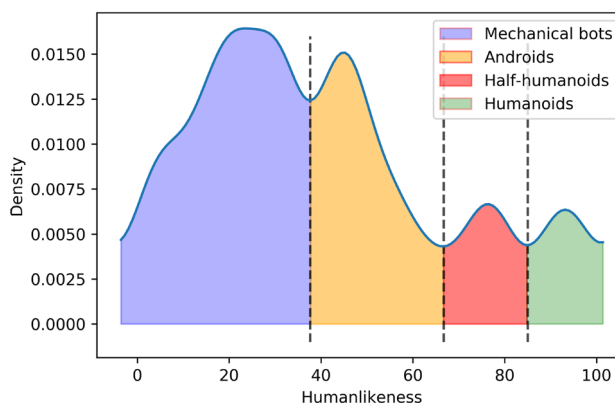
that people may treat smaller robots as if they were playable or related to fun (see [60,61]). Also, developed on the basis of the intuition that bigger robots can be seen as stronger and more threatening to people, H3B states that the decreased sentiment for bigger robots is related to perceived threateningness.

To test these explanations I defined two additional indices: ‘playfulness index’ and ‘threateningness index’, analogously to previous uncanny indices. I took the synonyms from the Merriam–Webster Online Thesaurus of the word ‘play’ in the following meaning: “activity engaged in to amuse oneself”, and of the word ‘threatening’ in the meaning: “involving potential loss or injury”. The synonyms used for the indices are presented in Table 5.

Subsequently, I then conducted mediation analysis, testing if playfulness and threateningness indices mediate the relationship between the height of robots and sentiment scores. The analysis was performed using R software [42] and the *mediation* package [56]. Using Baron and Kenny’s [2] procedure, I tested the influence of height (independent variable) on the playfulness and threateningness indices (mediators) separately. The model for playfulness was significant ( $F(31, 1) = 6.4, p = 0.017$ ), and the model for threateningness was not significant ( $F(31, 1) = 2.9, p = 0.1$ ). Therefore, I did not find evidence to support the hypothesis that threateningness is a mediator of the height and sentiment relationship. Next, I tested the combined influence of height and playfulness on sentiment. The effect of size on sentiment became no longer significant ( $t = -1.8, p = 0.09$ ), while the effect of playfulness remained significant ( $t = 3.3, p = 0.003$ ). Mediation schema with coefficients is presented in Fig. 9. The bootstrapping test with 10,000 simulations showed that the mediation was significant ( $ACME = -0.0018, CI [-0.0034, 0.0], p = 0.044$ ). Therefore, playfulness was found to be a significant mediator of the height and sentiment relationship. H3A is supported and smaller robots are perceived as more playable (as toys). H3B is rejected and it cannot be confirmed that people perceive bigger robots as more threatening.



**Fig. 9** Regression coefficients for the relationship between height and sentiment score as mediated by playfulness. The regression coefficient between height and sentiment, controlling for playfulness, is in the parentheses. \*  $p < .05$ , \*\*\*  $p < 0.001$



**Fig. 10** The kernel density estimate plot with distinguished groups

### 3.5 Exploratory Analysis

The following analysis is conducted in order to define the most suitable words for measuring self-reported attitudes toward robots. I grouped robots with similar humanlikeness scores (from ABOT database) to examine how people describe them. I used the kernel density estimation in order to distinguish groups within humanlikeness scale (gaussian kernel, bandwidth equal to 5)<sup>12</sup>. Four groups were distinguished: (1) mechanical bots – low humanlike robots with mechanistic surface, and low to medium humanlike facial features and body [humanlikeness score: 0–37]; (2) androids— medium humanlike robots with facial and bodily features, but with low humanlike surface [humanlikeness score: 37–67]; (3) half-humanoids—humanlike robots with surface, and facial features resembling humans, but without entirely humanlike body [humanlikeness score: 67–85]; and (4) humanoids— highly humanlike robots with humanlike surface, facial and bodily features [humanlikeness score: 85–100]. The kernel density estimate is presented in Fig. 10. Table 6 shows means

<sup>12</sup> Usage of the optimal bandwidth calculated according to Silverman’s rule of thumb ( $bw = 11.8$ ) resulted in one data cluster, therefore I had been decreasing the bandwidth until 4 groups were distinguished, as four is the minimal number of humanlikeness groups to describe the uncanny valley [25].

**Table 6** Average humanlikeness and its subscales scores for distinguished groups, and the robots classified within each group

Group	Hum.	Sur.	Face	Body	Robots in the group
1. Mechanical bots	19.6	0.06	0.47	0.44	Zenbo, Durus, Misty, Atlas, Ollie, Jibo, Cozmo, Cassie, Pr2, Walker, Loomo, Mip, Hitchbot, Handle (later), Spot, Handle (earier), Vector
2. Androids	49.3	0.18	0.81	0.9	Valkyrie, Pepper, Asimo, Nao, Nexi, Robonaut, Icub, Cb2, Kojiro, Qrio
3. Half-humanoids	76.3	0.77	0.99	0.37	Sophia, Bina48, Han
4. Humanoids	93.1	0.97	0.99	0.97	Jiajia, Erica, Nadine

of humanlikeness and subscales scores and the group assignment for particular robots.

In order to identify the language which people use to describe robots from the humanlikeness spectrum, I counted adjectives used for previously distinguished groups of robots. Adjectives are usually used for the description of the features and expressions of opinion in language [13,23], and are therefore good targets for attributes retrieval. Adjectives were counted for each robot and then normalized to the number of total words in the robot corpus in order to avoid enlarging the contribution of robots with more comments. For each group, I then counted the arithmetic mean of the frequencies of the words (due to a different number of robots in groups). Then, in order to identify adjectives that occur with unusual frequency in a given group, I counted scores for words according to the following equation:

$$score = \frac{f_w^2}{f_t}$$

where  $f_w$  is the word frequency within the group, and  $f_t$  is the word frequency in all groups. Such an estimation emphasizes words that are relatively frequent in a group in relation to other groups. The top 15 sorted adjectives with the highest score for each group are presented in Table 7.

Whereas for *mechanical bots* and *androids* groups it is hard to indicate any specific words (although the ones I identified all seem to be positively-valenced); for *half-humanoids* and *humanoids* it strikes that the words are related to the uncanny valley, i.e., ‘scary’ and ‘creepy’; and also to the artificial-real dimension, i.e., ‘human’, ‘real’, ‘fake’, ‘live’, ‘realistic’, ‘robotic’, ‘artificial’ and ‘android’. An extended list of relatively frequent adjectives is added to Supplementary Materials.

### 3.6 Uncanny Valley Awareness

I also wanted to test the awareness of UVH among commenters explicitly. I counted the frequencies of the ‘uncanny

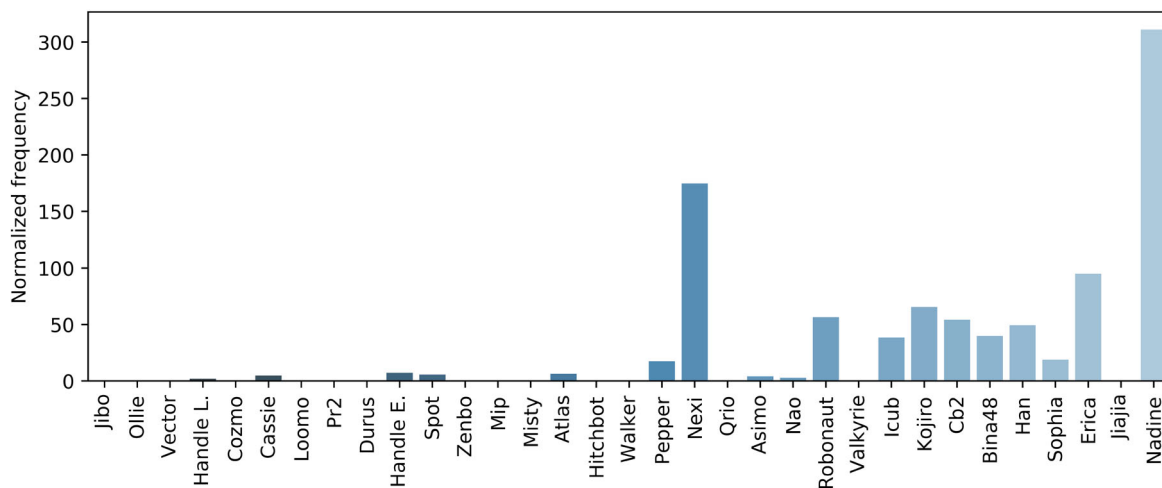
**Table 7** The most frequent words relative to other groups

Mech. bots	Androids	Half-humanoids	Humanoids
Good	Human	Human	Human
Cool	Cool	Nuclear	Real
Awesome	Little	Real	Beautiful
Much	Good	Fake	Japanese
Great	Cute	Creepy	Chinese
Cute	Better	Scary	Robotic
New	Much	First	Female
Best	New	Sure	Realistic
Little	Awesome	Old	Artificial
Better	Great	Stupid	Scary
Next	Able	Last	Android
Happy	Real	God	Young
Big	Advanced	Live	Sexual
First	Amazing	Weird	Bad
Flat	Japanese	Realistic	Hot

valley’ term appearing in comments. For several robots, there were no occurrences of the term, meaning that the chi-square test could not be used to statistically test differences. I present the normalized frequency in Fig. 11. The plot shows that the occurrence of the ‘uncanny valley’ term is frequent for more humanlike robots, whereas its occurrence for other robots is none or near to none (with the exception of the Nexi robot). The public seems to be aware, at least to some extent, of the existence of the uncanny valley.

## 4 Discussion

The aims of this paper were as follows: to test the shape of the relationship between robot humanlikeness and sentiment scores, to examine which of the variables (eeriness, pleasantness, and attractiveness) are related to humanlikeness, to



**Fig. 11** Normalized frequency of the ‘uncanny valley’ term occurrence; per 1000 comments. Robots are ordered by the humanlikeness score

test the impact of robot size on sentiment, and to characterize the specific, emotional words expressed toward robots. The study focused on providing ecologically-valid results about people’s emotional reactions toward robots. The acquisition of comments from the YouTube video-sharing platform allowed relatively natural utterances, not affected by experimental conditions, to be examined.

The analysis of robot-related comments supports the presence of a specific attitude toward very humanlike robots called the uncanny valley. The results show that people use words relating to the concept of eeriness to describe very humanlike robots. Given the large sample of data (224,544 comments for 33 robots), this is strong evidence, that the uncanny valley is a real issue and doubts of its existence (e.g., [4,25]) are not valid. Emotions manifested in UVH are limited to eeriness. The shape of general humanlikeness and sentiment is linear – more humanlike robots elicit more negative sentiment. One of the subscales of humanlikeness – Facial Features – shows a non-linear relationship with eeriness and sentiment. The attractiveness, related to mate selection, is one of the explanations of the uncanny valley (e.g., [8,32]). My results show no relationship between attractiveness and humanlikeness, and do not support this explanation. Additionally, the study shows that the size of robots can influence the general emotions toward them (mediated by the perception of smaller robots as designed for play), which is in line with [61].

#### 4.1 Shape of the Uncanny Valley

Mori [35] hypothesized the existence of a non-linear relationship between humanlikeness and affinity level. However, empirical studies advocate that affinity increases linearly across increasing humanlikeness [8,25]. The results of the YouTube comments analysis also support the hypothesis

that the relationship between humanlikeness and emotional valence (positive vs. negative) is linear, but not in the direction proposed by Kätsyri et al. [25]. According to the analysis, as humanlikeness increases, sentiment decreases. A reverse relationship seems to exist for eeriness—as humanlikeness increases, perceived eeriness increases.

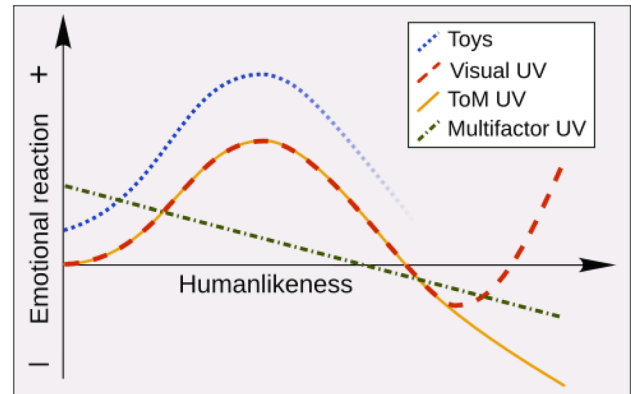
As for factors that underlie the humanlikeness (according to Phillips et al. [40]), only the Body-Manipulators subscale (presence of hands, arms, torso, fingers, and legs) does not influence the sentiment score (nor the eeriness). The impact of this subscale for UVH is limited, which is interesting, as the Body-Manipulators subscale was previously found to be the greatest contributor to the humanlikeness of robots [40]. It seems that Surface Look (presence of eyelashes, head hair, skin, genderedness, nose, eyebrows, apparel) and Facial Features (presence of a face, eyes, head, mouth) are the most important humanlikeness subscales for the UVH. There was a linear relationship between Surface Look and sentiment, whereby higher levels of Surface Look were associated with more negative sentiment. With respect to eeriness, there was a positive relationship with Surface Look up to a certain degree, but at the highest levels of Surface Look the pattern reversed and eeriness perceptions decreased. The Facial Features subscale shows a sinusoidal pattern both for sentiment and eeriness—a very high score on this scale seems to greatly decrease sentiment and increase eeriness. This relationship seems to resemble the characteristic dip of UVH. The dimension of Facial Features reflects people’s expectations that robots interact socially and effectively communicate with humans [40], therefore this is yet another argument for the involvement of social thinking in the uncanny valley effect [34,58].

Mathur and Reichling [34] showed a cubic relationship between mechano-humanness and likability after conducting an experimental study. Because they used images of

robot faces as stimuli, one may have suspected that the relationship would be similar to the relationship between Facial Features and sentiment in this study. However, cubic functions were mirrored (in [34], the cubic function approaches positive infinity, and in my study the function approaches negative infinity). Mathur and Reichling [34] asked subjects to estimate friendliness versus creepiness of possible interactions with a robot after a static image display. This stands in contrast to motion picture stimuli from my study. The method of stimuli presentation (images vs. movies) influence the elicitation of emotions [12] and this may be the reason for the different shapes of the obtained models. Possibly, people watching robot movies are able to judge the social behavior of robots, attributing agency, and experience (see [16]) far better than while watching images. The usage of static images in Mathur and Reichling [34] puts emphasis on visual aspects of robots, thus variable of static images may be more similar to the influence of surface look. This is in line with the similarity between results of eeriness index and Surface Look subscale relationship (see Fig. 6) and mechano-humanness and likability relationship from Mathur and Reichling [34]. Many of the uncanny valley studies focus mainly on visual aspects of robots, but the results of Stein and Ohler [50] showed that interacting characters may be seen as more eerie, depending on the beliefs of observers—if they think that characters are controlled by artificial intelligence, they assess characters as more eerie than when thinking that they are controlled by a human. This suggests that the Theory of Mind (e.g., [11]) factor may be involved in the uncanny valley effect. Mori [35] proposed that the movement of robots exaggerates the eeriness, but this has not been confirmed [41]. Presumably, this hypothesis should be modified, i.e., not the simple movement (as in the study of [41], which was a simple door-knocking movement), but the complex movement of robots, which is perceived as specific behavior, exaggerates or even changes the effect. This would suggest the necessity of analyzing the impact of robots' behavior in the context of UVH.

Considering the comparison of models and the plots of relationship, it seems that the rightmost part of Mori's plot (the “leave” from the valley of the most humanlike robots) is questionable. Under conditions taking into account the mind attribution, the relationship resembles a cliff (as suggested by [3]) rather than a valley. Either the shape of UVH should be reconsidered, or perhaps it is impossible to determine the real shape of the uncanny valley as there are no real robots that are indistinguishable or nearly indistinguishable from human beings, due to the current state of technology.

Based on the obtained results and the above-mentioned literature, a revised version of the uncanny valley plot and its modifications is considered in Fig. 12.



**Fig. 12** Modifications of UV relationship; Toys—agents perceived as toys, visual UV—relationship between humanlikeness based on visual aspects and sentiment, ToM UV—relationship between humanlikeness based on attribution of human mind and sentiment, multifactor UV—relationship between multifactorial humanlikeness and sentiment. Detailed discussion in the text

## 4.2 Emotions Describing UVH

The regression analysis showed that eeriness, pleasantness, and attractiveness were not equally related to humanlikeness (H2), and in fact, only the eeriness index is associated with humanlikeness. While controlling eeriness, pleasantness (defined by the property of “giving pleasure or contentment to the mind or senses”), and attractiveness (defined by the property of “being very pleasing to look at”<sup>13</sup>) did not emerge as significant variables for explaining the uncanny valley effect.

The exploratory analysis of adjectives reflects the attitudes evident in the regression analysis. For less humanlike groups (mechanical bots and androids), relatively the most frequent adjectives were positive or neutral and not specific. For the most humanlike robots, relatively the most frequent adjectives are related to the perception of eeriness, i.e., ‘scary’ and ‘creepy’, and also for artificial-real dimension, i.e., ‘human’, ‘real’, ‘fake’, ‘live’, ‘realistic’, ‘robotic’, ‘artificial’ and ‘android’. This means that the uncanny valley feelings, as well as humanlikeness itself, are among the most discussed topics in the comments specific to very humanlike robots. The emotional adjectives from the list seem to be the most suitable words for measuring a self-reported decrease in affinity related to the uncanny valley. This addresses the suggestion of Kätsyri et al. [25] about the necessity of such empirical studies. For humanoids, two distinguished words (‘sexual’ and ‘hot’) may be related to robots’ sexualization phenomenon identified by Strait et al. [52]. This observation supports their claim that objectification of female-gendered robots is a real issue.

<sup>13</sup> Both definitions are taken from <https://www.merriam-webster.com/thesaurus/> [Accessed: 07.12.2020].

It is also worth mentioning that some people are aware of the uncanny valley effect, which is shown in Fig. 11, and this topic is popular among the internet community. This part of the commenters either know the definition of UVH or implicitly understand it, because occurrence of the term is limited to humanlike robots (not generally to robots). The case of Nexi robot, which had relatively more mentions of the ‘uncanny valley’ term than robots with similar humanlikeness, suggests the implicit understanding of the phenomenon. The Nexi robot has highly developed facial expressions, which also suggests the importance of mind attribution for UVH. Perhaps, in future experiments, the participants’ awareness of the uncanny valley should be controlled in order not to allow the results of self-report experiments to be biased by implicit knowledge.

### 4.3 Impact of Robots’ Size

The results show a significant relationship between the height of robots and sentiment scores. Additionally, perception of robots as more playable (as toys) is a mediator of this relationship.

Mäkäräinen et al. [33] came up with the concept of ‘funcanny valley’, i.e., artificial characters may be seen in a funny way, regardless of their uncanniness. They conducted a study using, among others, exaggerated smiling characters which elicited a positive reaction, despite their increasing strangeness. They suggested that the negative affective reaction described by the uncanny valley concept could, in some cases, evoke the sensation of amusement, funniness, and humorousness. Although their study had a different methodology and focused on human characters, the operationalization of the funcanny valley concept to my results may identify the factor of size as a variable in the funcanny valley. However, this interpretation does not explain the results of Mäkäräinen et al. [33], and further analyses are needed to identify how broad this concept is.

Perhaps the perception of artificial/robotic characters as created for amusement masks the uncanny effect, which may explain some differences in the assessments of uncanny characters.

### 4.4 Limitations

The study presented in this paper has some limitations that should be highlighted. Firstly, demographic information of commenters is not available to acquire on YouTube. Therefore, it is not clear if the acquired data is representative for the population. The sample selection might be biased by the YouTube algorithm, which may recommend watching robot videos to people interested in the topic. As a result, people who watch robot videos may be exposed to more robot videos. This is a valid issue because seeing more films

portraying robots tends to be associated with more positive attitudes toward robots [44]. Also, not all users are willing to put a comment under videos. Type of personality or attitudes toward internet media may influence whether or not someone shares their opinion online (see [38]). However, one may notice that type of personality or prior experience in research participation may influence the willingness to participate in laboratory studies either (see [46]). Some people may not want to participate in scientific studies but are willing to express their opinion on the Internet. It has been shown that internet comments analysis may provide valuable information about attitudes and can help to understand human behavior (e.g., [5,24]), therefore such an analysis, despite its weakness, may be beneficial for HRI research.

Additionally, the findings are based on English comments searched for the US region. Given that cultural factors might influence attitudes toward social robots and the way we respond to them [28], the findings should be interpreted with caution due to the potential generalisability issue. It would be valuable to conduct a similar analysis for other languages and regions in the future.

Furthermore, the context and narratives in which robots are presented may have an impact on the viewer’s sentiment. For example, the word ‘eerie’ may not always reflect the sentiment toward the robot but could refer to other things presented in the video. I utilized a few means to decrease the possibility of such confounds. Firstly, I limited video search only to short videos below 4 min. The longer video is, the greater possibility that it will contain unwanted narratives or other unexpected content. Secondly, I excluded responses to the main comments (sub-comments), as they may explore side topics. Thirdly, I used multiple numbers of videos for each robot (45.9 on average, see the “Appendix” for individual numbers), therefore the effect of video context presumably has been averaged.

Although the initial number of robots prepared for the analysis was big (246), after filtering the sample size decreased to 33. This number is derived from the popularity of robots on the Internet and possibly, as robots are seen as an increasingly interesting topic, it will become possible to conduct similar analyses with a bigger sample in the future. The cut-off number of comments for robots (more than 200) was a trade-off between preserving the initial sample size and not leaving robots with too few comments for unbiased analysis. Small corpora (with fewer than 200 comments) may show random results, due to drift to various topics of movies.

In the analysis of sentiment and emotional indicators, negative forms of phrases have not been taken into account. During the preparation of the analysis, I made a few tries with negation detection in YouTube comments—I tried to negate all the words in a sentence when ‘not’ occurs. However, the lack of punctuation in many comments caused problems. E.g., the algorithm negated all the words in comments with

multiple sentences without punctuation, which distorted the results. As Heerschop et al. [19] showed that simple inversion of the polarity of sentiment when negation occurs had a marginal effect on the improvement of performance (even for more structured text than internet comments), I used the conventional method of frequency counting with *AFINN* package without negation handling.

The values of  $R^2$  in models analysis (Sects. 3.1, 3.2) are relatively low (besides the relationships with Facial Features), which may be justified by the various topics of analyzed movies. However, a large number of comments should compensate for this randomness and reveal underlying attitudes toward robots.

#### 4.5 Future Work

The YouTube comments seem to be a rich source of information regarding attitudes toward robots. Despite its limitations,

future HRI studies analyzing internet comments may provide more explanatory insights due to the advantages of ecological validity. Besides the possible sentiment analysis of comments in languages other than English, deeper semantic analyses may provide valuable information about what influences the acceptance of robots. Also, methods of Natural Language Processing may help understand the causes of the uncanny valley.

## Appendix

See Table 8.

**Table 8** List of all analyzed robots with number of comments and videos, means and standard deviations for sentiment, eeriness, pleasantness, attractiveness, humanlikeness, Surface Look, Facial Features, Body Manipulators scores, and height [cm]

Name	Com. no.	Vid. no.	M sent.	SD sent.	Eerie.	Pleas.	Attr.	HL	Surf.	Face	BM	Height
Asimo	7090	167	0.6	3.01	1.2	5.7	7.8	45.4	0.0	0.5	1.0	130.0
Atlas	85507	134	-0.09	3.18	1.4	6.1	6.6	31.0	0.1	0.4	0.6	188.0
Bina48	751	5	-0.12	3.06	8.0	6.3	9.5	73.0	0.9	1.0	0.0	44.2
Cassie	2082	15	0.6	2.69	1.9	6.0	6.2	14.9	0.0	0.0	0.2	100.0
Cb2	552	11	-0.22	2.41	9.7	2.9	6.5	64.5	0.3	1.0	1.0	130.0
Cozmo	12588	167	1.7	3.22	0.3	7.7	15.0	13.9	0.1	0.8	0.2	6.4
Durus	219	2	0.44	3.0	1.2	9.0	9.6	21.0	0.0	0.2	0.4	180.0
Erica	736	11	0.49	3.66	3.6	7.2	9.3	89.6	0.9	1.0	0.9	166.0
Han	202	7	0.29	3.38	8.1	4.0	5.8	77.0	0.7	1.0	0.3	78.3
Handle E.	9658	25	0.18	3.0	2.4	6.1	6.4	21.2	0.0	0.2	0.6	198.0
Handle L.	5197	9	0.36	2.83	2.2	6.6	7.9	8.8	0.0	0.0	0.6	200.0
Hitchbot	3466	19	0.36	3.51	0.4	8.2	6.8	31.6	0.2	0.9	0.9	120.0
Icub	520	46	0.69	3.2	3.0	8.1	7.4	55.4	0.2	0.8	1.0	104.0
Jiajia	308	13	0.18	3.01	2.4	4.7	10.6	93.3	1.0	1.0	1.0	160.0
Nao	3609	19	0.61	2.9	1.4	6.4	9.9	45.9	0.0	0.9	1.0	57.4
Nexi	858	3	0.07	3.47	11.6	9.0	8.7	42.3	0.3	1.0	0.6	122.0
Ollie	2653	39	0.76	2.55	0.2	6.3	7.0	3.6	0.0	0.1	0.1	12.0
Pepper	1140	62	0.9	3.03	2.3	5.3	11.2	40.5	0.2	1.0	0.9	120.0
Pr2	269	17	1.45	2.87	0.0	9.4	9.9	20.0	0.1	0.7	0.8	149.0
Qrio	297	18	0.8	3.52	0.6	2.7	6.4	44.1	0.1	0.7	0.8	60.0
Robonaut	354	37	0.71	3.41	1.6	5.7	5.9	47.0	0.1	0.5	1.0	101.6
Sophia	7925	89	-0.04	3.32	5.8	5.7	6.6	78.9	0.7	1.0	0.8	167.0
Spot	68263	93	0.31	4.84	3.4	6.6	8.8	24.3	0.0	0.0	0.4	84.0
Valkyrie	839	36	1.02	3.18	0.3	6.7	5.3	50.6	0.2	0.8	1.0	190.0
Vector	6373	241	0.9	2.94	0.7	6.3	10.8	6.1	0.0	0.8	0.2	6.1
Walker	287	11	1.16	3.1	2.7	7.3	7.3	33.9	0.0	0.5	0.3	145.0
Zenbo	320	28	0.82	2.97	4.0	7.9	11.9	24.9	0.2	0.9	0.1	62.0

**Acknowledgements** The author would like to give his thanks to Paweł Łupkowski, Michał Wyrwa, and Maciej Raś for comments on a draft of this paper.

**Data availability** The scores for analyzed robots are presented in the Appendix. The Supplementary Materials are available at <https://osf.io/cyvmx>. The additional datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The author declare that he has no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmed K, El Tazi N, Hossny AH (2015) Sentiment analysis over social networks: an overview. In: 2015 IEEE international conference on systems, man, and cybernetics. IEEE, pp 2174–2179
- Baron RM, Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 51(6):1173
- Bartneck C, Kanda T, Ishiguro H, Hagita N (2007) Is the uncanny valley an uncanny cliff? In: Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on, pp 368–373
- Bartneck C, Kanda T, Ishiguro H, Hagita N (2009) My robotic doppelgänger—a critical look at the uncanny valley. In: RO-MAN 2009—The 18th IEEE international symposium on robot and human interactive communication, pp 269–276. <https://doi.org/10.1109/ROMAN.2009.5326351>
- Beigi G, Hu X, Maciejewski R, Liu H (2016) An overview of sentiment analysis in social media and its applications in disaster relief. In: Sentiment analysis and ontology engineering, Springer, pp 313–340
- Benjamini Y, Krieger AM, Yekutieli D (2006) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3):491–507
- Burleigh TJ, Schoenherr JR (2015) A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? *Front Psychol* 5:1488
- Burleigh TJ, Schoenherr JR, Lacroix GL (2013) Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Comput Hum Behav* 29(3):759–771
- Burnham KP, Anderson DR (2002) A practical information-theoretic approach. Model selection and multimodel inference, 2nd edn. Springer, New York, p 2
- Cheetham M, Wu L, Pauli P, Jancke L (2015) Arousal, valence, and the uncanny valley: Psychophysiological and self-report findings. *Front Psychol* 6:981
- Dennett DC (1971) Intentional systems. *J Philos* 68(4):87–106
- Detenber BH, Simons RF, Bennett GG Jr (1998) Roll ‘em!: the effects of picture motion on emotional responses. *J Broadcast Electron Media* 42(1):113–127
- Fei G, Liu B, Hsu M, Castellanos M, Ghosh R (2012) A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In: Proceedings of COLING 2012: Posters, pp 309–318
- Fernández-Berrocal P, Extremera N (2006) Emotional intelligence and emotional reactivity and recovery in laboratory context. *Psicothema* 18:72–78
- Friedman B, Kahn Jr PH, Hagman J (2003) Hardware companions? What online aibo discussion forums reveal about the human–robotic relationship. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp 273–280
- Gray K, Wegner DM (2012) Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125(1):125–130
- Guzman E, Azócar D, Li Y (2014) Sentiment analysis of commit comments in github: an empirical study. In: Proceedings of the 11th working conference on mining software repositories, pp 352–355
- Hajar M et al (2016) Using youtube comments for text-based emotion recognition. *Procedia Comput Sci* 83:292–299
- Heerschop B, van Iterson P, Hogenboom A, Frasinca F, Kaymak U (2011) Analyzing sentiment in a large set of web data while accounting for negation. In: Advances in intelligent web mastering, vol 3, Springer, pp 195–205
- Ho CC, MacDorman KF (2010) Revisiting the uncanny valley theory: developing and validating an alternative to the godspeed indices. *Comput Hum Behav* 26(6):1508–1518
- Ho CC, MacDorman KF, Pramono ZD (2008) Human emotion and the uncanny valley: a glm, mds, and isomap analysis of robot video ratings. In: 2008 3rd ACM/IEEE international conference on human–robot interaction (HRI), pp 169–176
- Hover QR, Velner E, Beelen T, Boon M, Truong KP (2021) Uncanny, sexy, and threatening robots: The online community's attitude to and perceptions of robots varying in humanlikeness and gender. In: Proceedings of the 2021 ACM/IEEE international conference on human–robot interaction, pp 119–128
- Hu M, Liu B (2004) Mining opinion features in customer reviews. In: Proceedings of 19th national conference on artificial intelligence (AAAI'04), vol 4, pp 755–760
- Jelodar H, Wang Y, Orji R, Huang S (2020) Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J Biomed Health Inform* 24(10):2733–2742
- Kätsyri J, Förger K, Mäkäräinen M, Takala T (2015) A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Front Psychol* 6:390
- Kätsyri J, de Gelder B, Takala T (2019) Virtual faces evoke only a weak uncanny valley effect: an empirical investigation with controlled virtual face images. *Perception* 48(10):968–991
- Kim SY, Schmitt BH, Thalmann NM (2019) Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Mark Lett* 30(1):1–12
- Lim V, Rooksby M, Cross ES (2021) Social robots on a global stage: establishing a role for culture during human–robot interaction. *Int J Soc Robot* 13(6):1307–1333
- Löffler D, Dörrenbächer J, Hassenzahl M (2020) The uncanny valley effect in zoomorphic robots: the u-shaped relation between animal likeness and likeability. In: Proceedings of the 2020

- ACM/IEEE international conference on human–robot interaction, pp 261–270
30. Łupkowski P, Rybka M, Dziejdzic D, Włodarczyk W (2019) The background context condition for the uncanny valley hypothesis. *Int J Soc Robot* 11(1):25–33
  31. MacDorman KF, Chattopadhyay D (2016) Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition* 146:190–205
  32. MacDorman KF, Green RD, Ho CC, Koch CT (2009) Too real for comfort? Uncanny responses to computer generated faces. *Comput Hum Behav* 25(3):695–710
  33. Mäkäräinen M, Kätsyri J, Förger K, Takala T (2015) The uncanny valley: a study of positive emotional reactions to strangeness. In: *Proceedings of the 19th international academic mindtrek conference*. ACM, pp 175–181
  34. Mathur MB, Reichling DB (2016) Navigating a social world with robot partners: a quantitative cartography of the uncanny valley. *Cognition* 146:22–32
  35. Mori M (1970) The uncanny valley. *Energy* 7(4):33–35
  36. Ngai EW, Tao SS, Moon KK (2015) Social media research: theories, constructs, and conceptual frameworks. *Int J Inf Manag* 35(1):33–44
  37. Nielsen FÅ (2011) A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: *Proceedings of the ESWC2011 workshop on making sense of microposts* (2011)
  38. Orchard LJ, Fullwood C (2010) Current perspectives on personality and internet use. *Soc Sci Comput Rev* 28(2):155–169
  39. Palomäki J, Kunnari A, Drosinou M, Koverola M, Lehtonen N, Halonen J, Repo M, Laakasuo M (2018) Evaluating the replicability of the uncanny valley effect. *Heliyon* 4(11):e00939
  40. Phillips E, Zhao X, Ullman D, Malle BF (2018) What is human-like? Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. In: *Proceedings of the 2018 ACM/IEEE international conference on human–robot interaction*, pp 105–113
  41. Piwek L, McKay LS, Pollick FE (2014) Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130(3):271–277
  42. R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
  43. Ratajczyk D, Jukiewicz M, Łupkowski P (2019) Evaluation of the uncanny valley hypothesis based on declared emotional response and psychophysiological reaction. *Bio Algorithms Med Syst*. <https://doi.org/10.1515/bams-2019-0008>
  44. Riek LD, Adams A, Robinson P (2011) Exposure to cinematic depictions of robots and attitudes towards them. In: *Proceedings of international conference on human–robot interaction, workshop on expectations and intuitive human–robot interaction*, Citeseer, vol 6
  45. Rosenthal-Von Der Pütten AM, Krämer NC (2014) How design characteristics of robots determine evaluation and uncanny valley related responses. *Comput Hum Behav* 36:422–439
  46. Saliba A, Ostojic P (2014) Personality and participation: who volunteers to participate in studies. *Psychology* 5, 230–243
  47. Seyama J, Nagayama RS (2007) The uncanny valley: effect of realism on the impression of artificial human faces. *Presence Teleoperators Virtual Environ* 16(4):337–351
  48. Shank DB, Graves C, Gott A, Gamez P, Rodriguez S (2019) Feeling our way to machine minds: people's emotions when perceiving mind in artificial intelligence. *Comput Hum Behav* 98:256–266. <https://doi.org/10.1016/j.chb.2019.04.001>
  49. Siersdorfer S, Chelaru S, Nejdil W, San Pedro J (2010) How useful are your comments? Analyzing and predicting youtube comments and comment ratings. In: *Proceedings of the 19th international conference on World wide web*. ACM, pp 891–900
  50. Stein JP, Ohler P (2017) Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160:43–50
  51. Strait M, Ramos AS, Contreras V, Garcia N (2018) Robots racialized in the likeness of marginalized social identities are subject to greater dehumanization than those racialized as white. In: *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, pp 452–457
  52. Strait MK, Aguilon C, Contreras V, Garcia N (2017) The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a “technology takeover”, and the unabashed sexualization of female-gendered robots. In: *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp 1418–1423
  53. Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 29(1):24–54
  54. Tay T, Low R, Loke H, Chua YL, Goh Y (2018) Uncanny valley: A preliminary study on the acceptance of Malaysian urban and rural population toward different types of robotic faces. In: *IOP conference series: materials science and engineering*, vol 344. IOP Publishing, p 012012. <https://doi.org/10.1088/1757-899x/344/1/012012>
  55. Thelwall M (2018) Social media analytics for youtube comments: potential and limitations. *Int J Soc Res Methodol* 21(3):303–316
  56. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2014) mediation: R package for causal mediation analysis. *J Stat Softw* 59(5):1–38
  57. Urgen BA, Li AX, Berka C, Kutas M, Ishiguro H, Saygin AP (2015) Predictive coding and the uncanny valley hypothesis: evidence from electrical brain activity. In: *Cognition: a bridge between robotics and interaction*, pp 15–21
  58. Urgen BA, Kutas M, Saygin AP (2018) Uncanny valley as a window into predictive processing in the social brain. *Neuropsychologia* 114:181–185
  59. Vlachos E, Tan ZH (2018) Public perception of android robots: indications from an analysis of youtube comments. In: *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp 1255–1260
  60. Wagner D, Billinghamurst M, Schmalstieg D (2006) How real should virtual characters be? In: *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*. ACM, p 57
  61. Wang I, Smith J, Ruiz J (2019) Exploring virtual agents for augmented reality. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. ACM, p 281
  62. Wang S, Lilienfeld SO, Roach P (2015) The uncanny valley: existence and explanations. *Rev Gen Psychol* 19(4):393–407
  63. Yu CE (2020) Humanlike robots as employees in the hotel industry: thematic content analysis of online reviews. *J Hosp Mark Manag* 29(1):22–38

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Dawid Ratajczyk** is a cognitive scientist and a PhD student at the Adam Mickiewicz University in Poznan, Poland. Besides a master's degree in Cognitive Science, he also received a bachelor's degree in Biology. His current research interest is focused on social robotics, human-computer interaction, virtual reality, and language processing.

## **A.2 Study II: The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen**



## The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen

Dawid Ratajczyk, Jakub Dakowski & Paweł Łupkowski

**To cite this article:** Dawid Ratajczyk, Jakub Dakowski & Paweł Łupkowski (2023): The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen, International Journal of Human-Computer Interaction, DOI: [10.1080/10447318.2023.2179216](https://doi.org/10.1080/10447318.2023.2179216)

**To link to this article:** <https://doi.org/10.1080/10447318.2023.2179216>



Published online: 22 Feb 2023.



Submit your article to this journal [↗](#)



Article views: 6



View related articles [↗](#)



View Crossmark data [↗](#)



# The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen

Dawid Ratajczyk<sup>a,b</sup> , Jakub Dakowski<sup>a</sup> , and Paweł Łupkowski<sup>a,b</sup> 

<sup>a</sup>Faculty of Psychology and Cognitive Science, Adam Mickiewicz University in Poznań, Poznań, Poland; <sup>b</sup>Reasoning Research Group, Adam Mickiewicz University in Poznań, Poznań, Poland

## ABSTRACT

The current study had two main goals—evaluation of the impact of robot stimuli presentation type (virtual reality vs. on-screen simulation) and investigation of the relation between essentialism beliefs and the uncanny valley effect. The experiment involved a virtual café in which participants confronted four characters ranging from robotic to humanlike. The results showed that robotic characters were rated more eerie than humanlike characters. The robotic character with moderate humanlikeness received the highest eeriness rating, which was interpreted by us as evidence of the uncanny valley effect. Contrary to previous research's suggestion, participants assessed the humanlikeness and eeriness of characters in the same manner in VR and on-screen. Our results show the importance of essentialism beliefs in attitudes toward the most eerie artificial agents. Strong beliefs in the uniqueness of human nature deepen the uncanny valley effect. Additionally, groups with and without previous knowledge of the uncanny valley effect did not differ in their assessments of variables related to this phenomenon.

## 1. Introduction

At the beginning of the 20th century, Jentsch (1906/2008) noticed that “the life-size machines that perform complicated tasks” may evoke uneasiness and uncanny feelings. He emphasized that the more sophisticated and genuine the mechanism is, the stronger the feeling it evokes. This idea was forged by Mori (1970) into the uncanny valley (UV), i.e., the relation between humanlikeness of artificial agents and evoked eerie feelings, with an increased eeriness for almost humanlike agents. This greatly influenced human-machine interaction studies. A study conducted by Mathur and Reichling (2016) showed that UV is not a minor effect influencing human perception but it is able to penetrate more deeply, modifying actual trust-related social behavior in human-robot interaction. The growing percentage of elderly people in society and advancement in robotics make the use of robots in elder care increasingly likely (Robinson et al., 2014; Sharkey & Sharkey, 2012). The application of robots would be possible in areas, such as assistance in daily tasks, monitoring the behavior and health, and also providing companionship (Sharkey & Sharkey, 2012). However, to properly introduce robots, it is necessary to acknowledge the social implications of their usage, such as emotional reactions to them.

UV effect influences social judgments not only in the case of robot interaction but also acceptance of virtual avatars and chatbots (Ciechanowski et al., 2019; Milcent et al., 2022; Shin et al., 2019). The results from UV related studies may be used in finding the proper visual design, facial

expressions, demeanor, conversational flow, and other factors for good user experience (Schwind et al., 2018; Shin et al., 2019; Skjuve et al., 2019).

Nowadays, even though artificial agents are already present in people's life [e.g., companion robots in facilities for older people (Bradwell et al., 2020) and conversational chatbots (Liebrecht & Hooijdonk, 2019)], there still are unresolved issues regarding the extent, the shape of the relationship, and causes of UV phenomenon. Furthermore, among research that aims to resolve these issues, the operationalization of variables and utilization of methodologies varies between studies (Diel et al., 2021), making it difficult to determine whether differences in research results are due to differences in methodologies or other factors.

The usage of distinct entities as stimuli (such as images or videos of existing robots, androids, computer-animated characters, humans, or other entities) is one of the most common practices in UV studies (Diel et al., 2021). Many studies focused on the appearance of artificial characters using static images as stimuli (e.g., Mathur & Reichling, 2016; Seyama & Nagayama, 2007; Steckenfinger & Ghazanfar, 2009) or 3D moving stimuli (e.g., Kätsyri et al., 2017; Piwek et al., 2014; Siebert et al., 2020). Virtual reality (VR) head-mounted displays are also being used more frequently (e.g., Stein et al., 2019; Stein & Ohler, 2017). As these methods differ in stimuli presentation and yield mixed results, understanding the impact of presentation types on the UV effect is crucial.

Złotowski et al. (2015) showed that repeated interactions with robots may influence attitudes toward them. Also, some research has suggested that participants' prior knowledge and beliefs may influence their attitudes toward artificial agents (see Piçarra, 2014; Ratajczyk, 2022). Therefore the valid question is if knowledge about the UV effect is associated with the emotional perception of agents. As knowledge about this phenomenon is increasing, it is important to know whether such knowledge may be a confounding factor in UV-related studies.

### 1.1. Presence and proximity of characters

The implementation of VR devices in research is a new trend in psychology, allowing advantages in experimental control, reproducibility, and ecological validity (Pan & Hamilton, 2018). For instance, it was shown that VR can be successfully used as a replacement for standard exposure therapy (e.g., virtual spiders as a replacement for real ones) in several phobias (Wechsler et al., 2019).

The term “presence” is strongly associated with VR. It is a state of the subjective sense of being physically in a computer mediated environment (the illusion of reality) (Sanchez-Vives & Slater, 2005; Schubert et al., 2001). In other words, presence is a subjective description of how convincing subjects find the virtual environment. The sense of presence can be separated from the ability of technology to immerse a user. As such, immersion is an objectively measurable aspect of the technology, and presence is a variable of a user's experience (Schubert et al., 2001). Jennett et al. (2008) view immersion differently. For them, immersion is the prosaic experience of being engaged with a videogame. It is connected to notions like “cognitive absorption” and “flow”, and concerns the state of the user playing a game or performing a task. That being said, different technologies can still differ in how strongly they immerse the user. While the difference will vary across users, it would be valid to ask if both the perceived presence and also immersion (defined as an engagement) play a role in UV perception.

As troubles with empirical repetitiveness of findings in UV research are observed (Diel et al., 2021; Kätsyri et al., 2015), we wanted to look at the proximity variable from a different perspective, as the VR set-up, unlike standard desktop experiments, allows users to be present in front of virtual agents. Therefore, in a safe experimental set-up, the effect of uncanniness may be reduced due to the barrier of the screen.

One of Mori's (1970) notions was that the UV effect originates as a form of instinct that protects us from proximal (as opposed to distal) sources of danger. An uncanny character closer to us may be seen as more dangerous (in terms of physical damage or/and transmitting diseases) and also more eerie. Therefore, we may suspect that the proximity of characters affects perceived eeriness.

Stein et al. (2019) presented a model of proximal and distal threat dimensions from artificial agents. In their model, the most proximal source of danger was physical harm, the

intermediate source was loss of jobs and resources, and the most distal source of danger was loss of human uniqueness. Their findings showed that the closeness of characters did indeed influence emotional evaluations.

Kim et al. (2012) found that virtual environments and on-screen displays evoked different emotional reactions within their psychological experiment. Contrary to this result, Weidner et al. (2017) found no differences in psychophysiological responses during driving simulations in a virtual environment and 2D screen.

Hepperle et al. (2020) reported differences in the UV effect between stimuli presentation, with characters being assessed as more humanlike and more eerie in VR compared to on-screen. However, a small sample of participants and questionable choice of characters,<sup>1</sup> advocates for broadening this issue.

### 1.2. Essentialism beliefs and social categories

From the point of view of psychological essentialism, people believe that objects have essences (underlying natures) that make them what they are (Demoulin et al., 2006). This applies to the perception of social categories (e.g., ethnicity, gender, walk of life), as people attribute them underlying characteristics. The social categories are perceived as having essences, and the degree to which people essentialize these categories may differ. Eyssel and Kuchenbrandt (2012) showed that robots are also subject to processes of social categorization. During interaction with artificial agents, they may seem more intrinsically similar (e.g., perceiving as “all the robots are the same”), which may cause attribution of characteristics to the whole group, ignoring superficial differences between agents. Also, some people may be more willing to accept robots as a part of human social categories.

On the basis of the idea that the way people see robots as a social category may impact attitudes toward them, Piçarra (2014) created the Belief in Human Nature Uniqueness scale (BHNU), and used it to suggest that people who perceive humans as a more unique group reveal more negative attitudes toward robots. Also, MacDorman and Entezari (2015) showed that the conception of a stronger distinction between humans and robots is associated with the eeriness and perceived warmth of robots. Gong (2008) showed that anthropomorphism of agents is associated with their social perception. The more anthropomorphic a computer agent is, the more it is perceived as social, competent, and trustworthy.

The social perception is an important factor in human-robot interaction. In the current study, we focused on the area of UV and examined the association between beliefs in human uniqueness and UV-related emotion, i.e., eeriness.

### 1.3. Research questions

The following research questions have been formulated based on the literature overview presented above.

### 1.3.1. Differences in uncanny valley variables between display types

As some UV research uses VR environments, it is valid to assess the influence of character presentation—particularly if it influences the assessment of variables related to the UV. Therefore, we decided to test whether people assess human-likeness and perceived eeriness differently in virtual reality and on computer desktop (RQ1, RQ2).

*RQ1. Do people assess the humanlikeness in VR and on-screen in the same way?*

To answer this question, we compared reports of human-likeness for four artificial humanoid agents or characters) using VR as experimental and computer screen as control group (between-subjects). In addition, we compared the humanlikeness of characters (within-subjects).

*RQ2. Does perceived eeriness differ for conditions and characters?*

We measured participants' evoked feelings of eeriness after encountering characters in a virtual café and compared it in a between-subject design. We also compared differences between characters (within-design).

### 1.3.2. Relation between immersion and eeriness

VR generally evokes higher immersion than 2D display, but the immersion may differ greatly between participants (see Jennett et al., 2008). The immersion questionnaire used in the study tests the engagement, focused attention, and emotional attitude toward the game/task. As participants may have various attitude and involvement into the scientific study (which may influence immersion), we decided to test whether immersion level is related to eeriness assessment (regardless of the method of stimuli presentation) (RQ3).

*RQ3. Is immersion level related to eeriness assessment?*

We measured participants' immersion level after the experiment to test for a correlation with eeriness.

### 1.3.3. Relation between beliefs and knowledge, and eeriness

Beliefs in human uniqueness are—among others—the manifestation of essentialism beliefs in social categorization. The beliefs in distinctiveness of artificial agents from human society may influence the emotions expressed toward these agents. Therefore we decided to test how such beliefs are associated with eeriness assessments (RQ4). Following the notion that prior knowledge regarding social robotics may influence attitudes toward artificial agents, we also decided to test whether *explicite* knowledge about the existence of UV effect is associated with the perceived eeriness of agents (RQ5).

*RQ4. Are essentialism beliefs related to eeriness assessments?*

We used the BHNU scale to measure beliefs in human nature uniqueness, and correlated the participants' results with eeriness. Additionally, we divided participants into

groups with low and high BHNU scores (based on the median) and compared their eeriness assessments.

*RQ5. Is UV knowledge related to different eeriness assessments?*

We compared reported eeriness between participants who did and did not know of the phenomenon.

## 2. Materials and methods

### 2.1. Participants

We recruited 94 participants (74 female) for a study assessing virtual humanoid agents. Participants were divided equally between the two conditions.

Participants were predominantly university students, mostly majoring in psychology and cognitive science. The mean age was 20.5 ( $SD = 2.2$ ; range = 17–28). All of them were rewarded with 25 PLN (about 5.50 USD) coupons to “Empik”—a popular Polish media store. Some of the undergraduates also received extra credit for participating. Ethical committee opinion was not required due to the non-interventional nature of the study.

### 2.2. Procedure

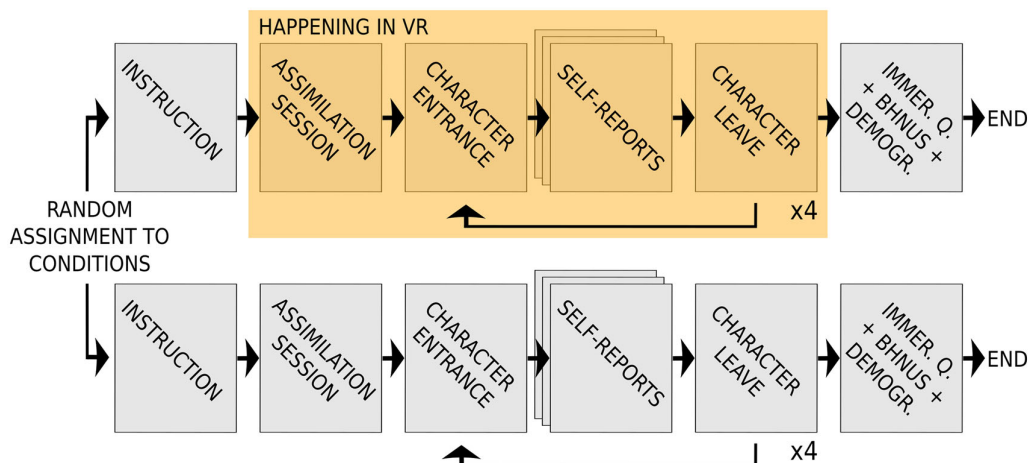
The participants were exposed to humanlike characters in the simulated environment created in the Unity game engine and asked to assess the humanlikeness and the eeriness of characters (see below). Half of the participants performed the task in VR using a head-mounted display and a VR controller, while the control group used a computer screen and a mouse (as presented in Figure 1). Therefore, the experimental difference was perceived *presence* influenced (evoked) by the type of display. After this, both groups were asked to fill out the Immersion Questionnaire (Jennett et al., 2008), the Belief in Human Nature Uniqueness scale (Piçarra, 2014), and metrics (all paper-pencil).

The participants were randomly assigned into groups. The examination started with the experimenter explaining the procedure of the experiment and obtaining informed consent from a participant. To minimize the influence of subjects' opinions on the results, they were told the study aimed to assess the qualities of the presented virtual characters. This was followed by a demonstration of the VR set and how it will be used during the study. The control group was given a brief explanation of how to mark their answers using a computer mouse.

Next, the participants were prepared for the study, and a short training phase took place. During this phase, the experimenters did communicate with the subjects to ease the learning process and make adjustments to their headsets.

The experimental phase started with a random character entering the room and waving to the participant. Then the participant was asked to answer five questions regarding humanlikeness and perceived eeriness. The answers were given using a VR controller in the experimental group and a mouse in the control group.

After that, the character left the room, and the procedure repeated until all four characters had been rated (see Figure 3). The participants were then asked to complete two



**Figure 1.** The experiment schema. The experimental condition used VR, while the control condition used a regular desktop monitor.



**Figure 2.** Participant in VR condition.

questionnaires and provide additional information about their age, gender, and knowledge of the UV effect. This was connected to their answers from the previous phase using a randomly generated key. Finally, participants were debriefed, and the goal of the study was revealed.

### 2.3. Experiment set-up

The study was performed in the Reasoning Research Group (RRG) laboratory at Adam Mickiewicz University in Poznań. In both groups, the participants were sat at a table during the experiment. In the control group, the computer screen and the mouse were placed on the table. For the VR condition, the table and chair positions corresponded to the position of the table and seat in the virtual café, keeping the same distances, to strengthen the presence. The VR set-up is presented in [Figure 2](#). There was no avatar of participants present in the virtual environment.

The experimental group used a VR headset HTC VIVE Cosmos with  $1440 \times 1700$  px per eye resolution ( $110^\circ$  field

of view). The control group used an LCD monitor with  $1920 \times 1080$  px resolution and 21.5" screen size.

### 2.4. Stimuli

The stimuli comprised four computer-generated characters from different levels of the humanlikeness spectrum (see [Figure 3](#)), implemented in a virtual café.<sup>2</sup> Characters were meant to represent four categories differing by humanlikeness as distinguished in previous studies: robots, androids, animated characters, and humans (Łupkowski et al., 2019; Ratajczyk et al., 2019). We used only male characters (for those that can be distinguished by gender) to not introduce an additional variable to the humanlikeness. All characters were animated in the same way (they walked into the room, waved to the participant, waited while the participant answered the questions, and left the room). Besides the experimental characters, there was a bartender at the far end of the room during the entire experiment to increase the reality of the environment. The same scene

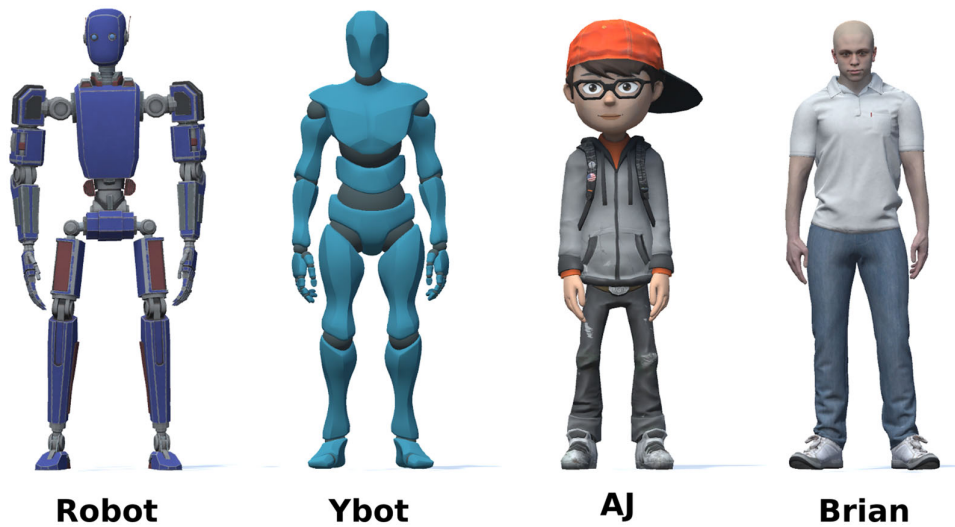


Figure 3. Characters used in the study ordered by humanlikeness evaluation.



Figure 4. Screenshot of the experiment in the on-screen (control) group. The original phrases were displayed in Polish.

was presented in both conditions: participants in VR sat in the café, whereas the control group watched the scene on the monitor screen. A screenshot of the experiment from the control group is presented in Figure 4. After the study, we checked whether the conditions were immersive to varying degrees. Analyses confirmed that conditions differed significantly in immersion level, with the VR condition evoking higher immersion ( $U = 652$ ,  $p = 0.001$ ,  $r = 0.40$ ).

## 2.5. Measures

All the questionnaires were in Polish.

### 2.5.1. Humanlikeness index

During the study, the participants were asked to assess the humanlikeness of the characters on a 5-point Likert scale. The same question was used in previous studies (Łupkowski et al., 2019; Ratajczyk et al., 2019). The results of this question were treated as the humanlikeness index.

1. How much does the presented model resemble a human? (Answers on a scale of 1–5, where (1) = Completely not humanlike, (2) = Rather not humanlike, (3) = It starts to look humanlike, (4) = Rather humanlike, (5) = Completely humanlike.)

### 2.5.2. Eeriness index

Following the method of Gray and Wegner (2012), the eeriness index was counted as a sum value of four intensity questions related to eeriness (Polish translations of “frightening,” “repulsive,” “strange,” “disturbing”<sup>3</sup>). Those terms were chosen to properly capture the idea of an UV dependent variable (eeriness) in Polish. Answers were incorporated on a scale of 1–4, where: (1)=Not at all, (2)=A little, (3)=Rather, (4)=Very much. Cronbach’s alpha for the eeriness questions indicated a good level of reliability ( $\alpha = 0.81$ ).

1. How frightening is this character?
2. How repulsive is this character?
3. How strange is this character?
4. How disturbing is this character?

### 2.5.3. Immersion questionnaire

We used a modified version of an immersion questionnaire (Jennett et al., 2008), adapted for Polish speakers by Strojny and Strojny (2014). As the questionnaire was primarily prepared for games, we used a modified version that was semantically suitable for utilization in an experiment (originally prepared by Kłosowski et al., 2019). A few questions related strictly to the game experience were deleted. The questionnaire consisted of 20 questions ( $\alpha = 0.83$ ) and is available in Supplementary Materials.<sup>4</sup>

### 2.5.4. BHNU scale

The Belief in Human Nature Uniqueness scale evaluates the extent to which people attribute uniquely human features (e.g., language, emotions, or consciousness) to robots (Piçarra, 2014; Pochwatko et al., 2015). Therefore, the results of this questionnaire indicate how one’s concept of the essence of human-being is limited and restricted to only humankind. Higher values indicate a stronger conviction in the uniqueness of human nature. It contains six questions ( $\alpha = 0.83$ ) and is available in Supplementary Materials.

### 2.5.5. Declared UV knowledge

To check for prior knowledge of the UV phenomenon, we explicitly asked participants: “Do you know the concept of the uncanny valley?” [with answers (1) = Yes, (2) = It’s hard to say, (3) = No].

## 3. Results

The analysis was performed in Python 3.8 and R software (R Core Team, 2021; RStudio Team, 2020). The threshold for statistical significance was set to  $p < 0.05$ . To reduce the probability of type I error, we used Benjamini–Hochberg correction when performing multiple comparisons.

### 3.1. Do people assess the humanlikeness in VR and on-screen in the same way?

We performed a nonparametric rank-based equivalent of the mixed ANOVA test due to unsatisfied assumptions for

parametric tests in the humanlikeness variable.<sup>5</sup> We used the ANOVA-type test proposed by Noguchi et al. (2012) implemented in R software as nparLD package (Noguchi et al., 2012). We performed a 2 (conditions: VR and on-screen)  $\times$  4 (characters: Robot, Ybot, AJ, and Brian) test with humanlikeness as the dependent variable. Neither the main effect for conditions,  $F(1, \infty) = 1.05$ ,  $p = 0.306$ , nor the interaction between conditions and characters,  $F(2.89, \infty) = 1.54$ ,  $p = 0.204$ , were significant. However, there was a significant main effect for characters,  $F(2.89, \infty) = 374.46$ ,  $p < 0.001$ , indicating that the effect of character on humanlikeness ratings was similar between the VR and on-screen conditions.

Pairwise Wilcoxon tests with Benjamini–Hochberg adjustment for multiple comparisons were performed as a *post hoc* analysis. Eta squared values were counted as effect size measures for *post hoc* tests. We observed significant differences between all characters.<sup>6</sup> Figure 5 shows the differences in humanlikeness across conditions.

### 3.2. Does the perceived eeriness differ for conditions and characters?

Eeriness data also did not satisfy assumptions for parametric tests.<sup>7</sup> Therefore, we performed a 2 (conditions: VR and on-screen)  $\times$  4 (characters: Robot, Ybot, AJ, and Brian) non-parametric rank-based ANOVA-type test with eeriness as the dependent variable (see Noguchi et al., 2012). Neither the main effect for conditions,  $F(1, \infty) = 0.80$ ,  $p = 0.368$ , nor the interaction between conditions and characters,  $F(2.56, \infty) = 1.67$ ,  $p = 0.179$ , were significant. However, a significant main effect was observed for characters,  $F(2.56, \infty) = 21.78$ ,  $p < 0.001$ , indicating that the differences in characters’ elicited eeriness were similar across the presentation method.

Pairwise Wilcoxon tests with Benjamini–Hochberg adjustment for multiple comparisons were performed as a *post hoc* analysis. Eta squared values were counted as effect size measures for *post hoc* tests. We observed significant differences between all but the two most humanlike characters

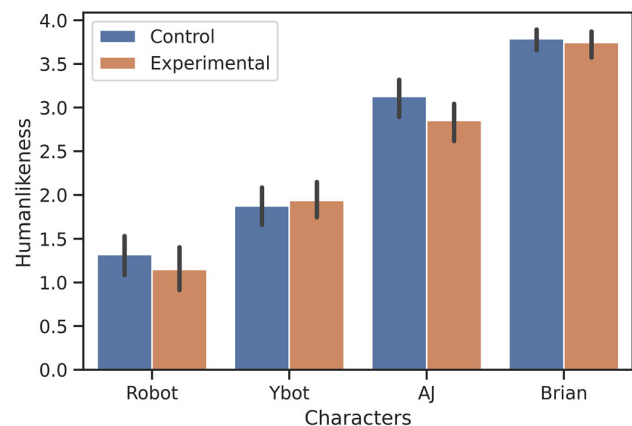


Figure 5. Differences in humanlikeness between characters. No significant difference between conditions was observed.

(AJ and Brian).<sup>8</sup> Figure 6 shows the differences in perceived eeriness across conditions.

### 3.3. Is immersion level related to eeriness assessment?

To examine whether immersion level is related to eeriness assessment, we tested Spearman's correlations between immersion questionnaire results and eeriness assessments for each character (four comparisons with Benjamini-Hochberg correction). The results were not significant for any characters: Robot:  $r_s = 0.11$ ,  $p = 0.618$ ; Ybot:  $r_s = 0.23$ ,  $p = 0.118$ ; AJ:  $r_s = 0.01$ ,  $p = 0.900$ ; Brian:  $r_s = 0.06$ ,  $p = 0.722$ . We therefore cannot say that differences in elicited eeriness were related to perceived immersion.

### 3.4. Are essentialism beliefs related to eeriness assessment?

To examine whether beliefs regarding human uniqueness are associated with eerie feelings toward humanlike characters we tested Spearman's correlations between BHNU scale results and eeriness assessments for each character (four comparisons with Benjamini-Hochberg adjustment). For Robot, AJ, and Brian the results were not significant (respectively:  $r_s = 0.12$ ,  $p = 0.327$ ;  $r_s = -0.16$ ,  $p = 0.225$ ;  $r_s = 0.008$ ,  $p = 0.938$ ). However, for the Ybot character the results were significant:  $r_s = 0.30$ ,  $p = 0.015$ . The correlation plot is presented in Figure 7.

We explored the obtained effect by dividing the participants into two groups based on their BHNU results. Participants with a BHNU level below or equal to 26 (the median) were labeled as "low BHNU" (48 people), whereas above 26 as "high BHNU" (46 people). We compared eeriness assessments between low and high BHNU groups (four comparisons with Benjamini-Hochberg adjustment). The Mann-Whitney test for independent groups was significant for Ybot ( $U = 1510$ ,  $p = 0.008$ ). The test showed no significant results for the remaining characters.<sup>9</sup> Figure 8 shows the differences in eeriness levels between low and

high BHNU groups. On the plot, we additionally presented polynomial curve fits based on mean eeriness for sub-groups. The results show that essentialism beliefs are related to the eeriness assessment, but only for the most eerie character.

### 3.5. Is UV knowledge related to different eeriness assessment?

To test whether knowledge of the UV effect is related to participants' eeriness assessments we analyzed responses to the question "Do you know the concept of the uncanny valley?" We dropped the group with the answer "It is hard to say," due to unknown status (8 people) and compared the remaining two groups ("yes" answer = 22 people, and "no" answer = 64 people). We tested for differences in eeriness assessment toward each character with Mann-Whitney (four comparisons with Benjamini-Hochberg adjustment). There were no significant differences between groups,<sup>10</sup> showing that knowledge of the concept of the UV did not alter eeriness assessments.

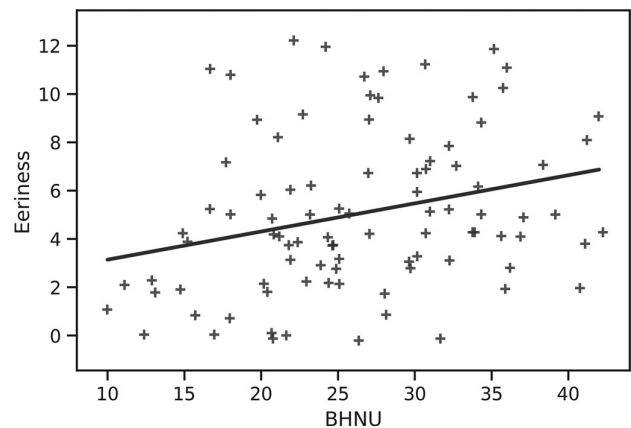


Figure 7. Spearman's correlation plot between beliefs in human uniqueness scale and eeriness assessments for Ybot. The points were jittered to avoid overplotting.

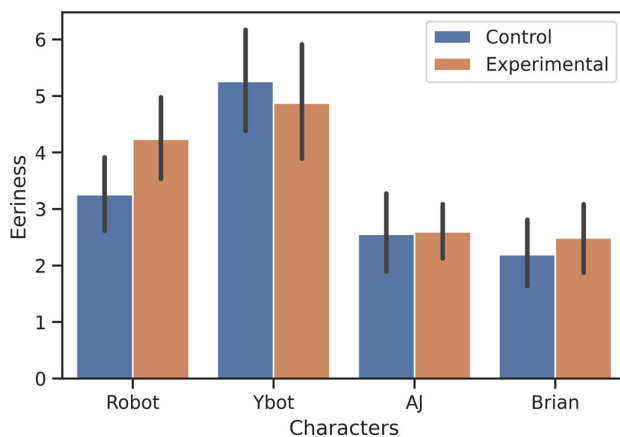


Figure 6. Differences in elicited eeriness between characters. No significant difference between conditions was observed.

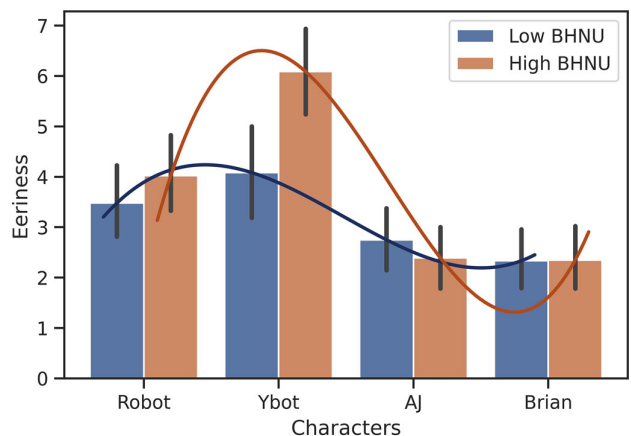


Figure 8. Differences in reported eeriness between low and high beliefs in human uniqueness scale groups; the presented curves are polynomial fits calculated on the basis of averaged eeriness.

## 4. Discussion

New technologies, such as the utilization of head-mounted displays to present in VR, offer new possibilities for human-machine interaction studies. The present study tested differences in the perception of artificial agents in VR and on-screen simulations. The experiment involved a virtual café in which participants confronted four characters ranging from robotic to humanlike. The results show that robotic characters were seen as more eerie than more humanlike characters. The highest eeriness rating was for the robotic character with moderate humanlikeness, which was interpreted by us as evidence of the UV effect. This seems to be in contrast with some previous studies (e.g., Yam et al., 2021), i.e., not all mechanical looking robots are less eerie than more humanlike. Tinwell et al. (2011) suggested that uncanniness may occur due to difficulties with emotion detection that hinder predictions about the behavior. The android figure in our study did not have facial expression, what could lead to more difficulties with mind attribution, increasing uncertainty, and making it harder to access intentions of the robot.

Contrary to suggestion from previous research (Hepperle et al., 2020), participants assessed the humanlikeness and eeriness of characters in VR and on-screen in the same manner. Our results show also the importance of essentialism beliefs in attitudes toward the most eerie artificial agents. Additionally, groups with and without knowledge of the UV effect did not differ in their assessments of variables related to this phenomenon.

### 4.1. Differences between VR and computer screen displays

We tried to answer the question if people assess the humanlikeness in VR and on-screen in the same way (RQ1) and also if perceived eeriness differs for conditions and characters (RQ2). Previous research has shown inconsistencies regarding users' differences in elicited emotions between VR and desktop displays (Kim et al., 2012; Weidner et al., 2017). Hepperle et al. (2020) explicitly compared the effect of these presentations on the UV, finding that people generally assessed characters as more humanlike and more eerie in VR. However, the size effect of differences was small<sup>11</sup> and based on small sample sizes (<15 participants per group). Furthermore, the stimuli fell within an excessively wide range of humanlikeness, including zombies and animalistic figures which are not suitable for measuring UV (Kätsyri et al., 2015), because zombies are deliberately designed to evoke negative emotions. As such, measures of eeriness could have been unrelated to humanlikeness. Additionally, their stimuli were static and isolated from any environment, which suggests that the study was focused solely on appearance.

Our results show no significant differences in assessment of eeriness and humanlikeness between VR and on-screen conditions. In our study, we used a simulation of encountering characters in café. Given the larger sample of

participants, embedding in the environment, and more appropriate selection of characters, we consider that the impression of presence in virtual environments do not contribute to differences in perception of humanlikeness and eeriness.

Although there are no differences in humanlikeness and eeriness assessments between VR and on-screen conditions, we still think that utilization of head-mounted displays may be beneficial for the research due to their increased ecological validity. The form of presentation does not influence the eeriness and humanlikeness of characters; however, participants' sense of realism may be stronger (see Diemer et al., 2015). As recent research suggests the involvement of mind perception in the UV phenomenon (Appel et al., 2020; Gray & Wegner, 2012; Stein & Ohler, 2017), it is important to involve the behavior of artificial agents in studies. The usage of a virtual environment seems to be a suitable intermediate stage between image studies and real robots studies, giving the benefits of more genuine interactions (see Wechsler et al., 2019), and relatively smaller costs (the expensiveness of real robot development).

### 4.2. Immersion and eeriness assessments

Jennett et al. (2008) proposed to consider immersion as related to the cognitive absorption, flow, emotional involvement, and engagement in a game or a task. We derived a research question whether immersion level is related to eeriness assessment (regardless of the method of stimuli presentation) (RQ3). The VR condition was indeed more immersive than computer-display condition, however our result did not show relation between reported immersion and eeriness. Jennett et al. (2008) stated also that immersion may be related to both the positive and negative emotions. Therefore people who perceive artificial humanoid characters as more positive or as more negative may all have a great impression of immersion. This issue needs more clarification in further studies.

### 4.3. Importance of essentialism beliefs

Essentialism is related to the perception of social groups as more or less entitative, natural, and unique (e.g., in terms of common goals, common fate, or similarity). Lay people who are strong essentialists perceive that group members very much resemble one another at a deep level (see Demoulin et al., 2006). Specifically, in the matter of human-machine interaction, robots, and other artificial agents may be seen as fundamentally different than humans. According to our results, social categorization has particular importance for the perception of the most uncanny characters. This result is in line with previous reports that a stronger belief in human nature uniqueness is associated with less favorable attitudes toward robots and a higher level of eeriness (MacDorman & Entezari, 2015; Piçarra, 2014; Złotowski et al., 2017). However, these studies did not report any involvement of humanlikeness,

i.e., that only characters with particular humanlikeness may influence different emotional reactions in believers in human nature uniqueness.

This result has two implications for human-machine interaction research, which will be discussed in turn. Firstly, it may be relevant to consider essentialism beliefs as a variable in the UV phenomenon. Beliefs in human uniqueness are strongly correlated with religiousness (Piçarra, 2014); however, religiousness has no direct influence on perception of artificial agents as more eerie (MacDorman & Entezari, 2015). It seems that beliefs in human uniqueness mediate the impact of religiousness on uncanny feelings toward artificial agents. Stronger beliefs in human uniqueness may be related to the threat of human identity in confrontation with artificial agents. Indeed, people who value social inclusion feel a stronger aversion to the autonomous AI agents (Stein et al., 2019).

UV research has focused on the category uncertainty explanation in recent years (e.g., Burleigh et al., 2013; MacDorman & Chattopadhyay, 2016; Palomäki et al., 2018). According to this explanation, the eerie sensation is caused by difficulties in the categorization of characters that lie between the robotic and human categories. Kätsyri et al. (2015), based on an extensive review, showed that the perception of characters ranging from artificial to human is indeed categorical. However, they did not find strong support for the hypothesis that ambiguous categorization induces negative affinity.

Reasonably, humanlikeness may not be related to perceptual categories, but to social categories, i.e., a certain degree of humanlikeness allows a viewer to attribute a human social category to the artificial agent. Such an agent may be threatening to the human social category. Characters closer to the boundary would therefore be more dangerous to the extension of the category, and thus to the identity of it. Results acquired in the current study are not sufficient to resolve this issue; however, such a rephrasing would explain why only the moderately humanlike robot was assessed as more eerie by participants who believed in strong human nature uniqueness.

The second implication for human-machine interaction research involves participant selection. If a large proportion of participants in a study happen to hold weak beliefs in human uniqueness, the obtained UV may become flattened, and the effect may be impossible to detect. This may be the reason for inconsistent results among UV studies. As students are commonly participants in scientific studies and also one of the most non-theistic groups (see Bullivant, 2008), this is a particularly relevant issue.

#### 4.4. Knowledge about the uncanny valley

Ratajczyk (2022) performed an analysis of YouTube comments regarding robots. The article reported that knowledge about the UV is present among lay people. It was important to examine whether knowledge about the effect may be related to a different assessment of eeriness, and thus should be controlled. Our results showed that there is no difference

between groups with and without knowledge about the UV effect. Thus, this issue requires further study.

#### 4.5. Limitations

The study has certain limitations. Firstly, the results regarding the association between beliefs in uniqueness of human nature and eeriness are not causal. Therefore we cannot say that beliefs in human nature uniqueness influence the emotional assessment of artificial agents. An additional correlated variable may be associated.

Secondly, we explicitly asked participants whether they knew the term “uncanny valley.” This does not involve implicit knowledge about the phenomenon. Perhaps a test of knowledge would be more informative and gain more reliable results.

Also, the participants in the experiment were mostly social science students of a similar age. We still do not know how much age influences attitudes toward robots (see Borau et al., 2021; Wasielewska & Łupkowski, 2021). As it is common practice to recruit students, we note that the generalization possibility of these results to the whole population may be limited.

#### Notes

1. They used a very broad range of humanlike characters, including animalistic and zombies, which was dissuaded by Kätsyri et al. (2015).
2. Characters were downloaded for free from [www.mixamo.com](http://www.mixamo.com) and [www.cgtrader.com](http://www.cgtrader.com) websites. The café model was bought from [www.turbosquid.com](http://www.turbosquid.com). All accessed 04.03.2022.
3. Polish translations: “przerazający”, “odpychający”, “dziwny”, “niepokojący”.
4. <https://osf.io/kr94u/>
5. None of the subgroups was distributed normally according to Shapiro-Wilk tests.
6. Robot vs. Ybot:  $W = 224$ ,  $p < 0.001$ ,  $\eta^2 = 0.15$ ; Robot vs. AJ:  $W = 0$ ,  $p < 0.001$ ,  $\eta^2 = 0.54$ ; Robot vs. Brian:  $W = 0$ ,  $p < 0.001$ ,  $\eta^2 = 0.77$ ; Ybot vs. AJ:  $W = 135$ ,  $p < 0.001$ ,  $\eta^2 = 0.35$ ; Ybot vs. Brian:  $W = 0$ ,  $p < 0.001$ ,  $\eta^2 = 0.71$ ; AJ vs. Brian:  $W = 53$ ,  $p < 0.001$ ,  $\eta^2 = 0.27$ .
7. See note 5 above.
8. Robot vs. Ybot:  $W = 918$ ,  $p < 0.001$ ,  $\eta^2 = 0.05$ ; Robot vs. AJ:  $W = 1076$ ,  $p < 0.001$ ,  $\eta^2 = 0.06$ ; Robot vs. Brian:  $W = 951$ ,  $p < 0.001$ ,  $\eta^2 = 0.08$ ; Ybot vs. AJ:  $W = 754$ ,  $p < 0.001$ ,  $\eta^2 = 0.17$ ; Ybot vs. Brian:  $W = 551$ ,  $p < 0.001$ ,  $\eta^2 = 0.20$ ; AJ vs. Brian:  $W = 1503$ ,  $p = 0.454$ ,  $\eta^2 = 0.003$ .
9. Robot:  $U = 1263$ ,  $p = 0.455$ ; AJ:  $U = 987$ ,  $p = 0.496$ ; Brian:  $U = 1099$ ,  $p = 0.969$ .
10. Robot:  $U = 703$ ,  $p = 0.992$ ; Ybot:  $U = 674$ ,  $p = 0.992$ ; AJ:  $U = 607$ ,  $p = 0.662$ ; Brian:  $U = 569.0$ ,  $p = 0.662$ .
11. They did not report effect sizes, however, we calculated them on the basis of reported tests results.

#### Disclosure statement

No potential conflict of interest was reported by the author(s).

#### Funding

This work was supported by the National Science Centre, Poland under Grant no. 2020/37/N/HS6/01372.

## ORCID

Dawid Ratajczyk  <http://orcid.org/0000-0002-8323-9935>

Jakub Dakowski  <http://orcid.org/0000-0002-7811-4580>

Paweł Łupkowski  <http://orcid.org/0000-0002-5335-2988>

## Data availability statement

The Supplementary Materials are available at <https://osf.io/kr94u/>. The additional datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## References

- Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior*, 102(January), 274–286. <https://doi.org/10.1016/j.chb.2019.07.031>
- Borau, S., Otterbring, T., Laporte, S., & Fosso Wamba, S. (2021). The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing*, 38(7), 1052–1068. <https://doi.org/10.1002/mar.21480>
- Bradwell, H. L., Winnington, R., Thill, S., & Jones, R. B. (2020). Longitudinal diary data: Six months real-world implementation of affordable companion robots for older people in supported living. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 148–150).
- Bullivant, S. (2008). Research note: Sociology and the study of atheism. *Journal of Contemporary Religion*, 23(3), 363–368. <https://doi.org/10.1080/13537900802373114>
- 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. <https://doi.org/10.1016/j.chb.2012.11.021>
- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92(March), 539–548. <https://doi.org/10.1016/j.future.2018.01.055>
- Demoulin, S., Leyens, J.-P., & Yzerbyt, V. (2006). Lay theories of essentialism. *Group Processes & Intergroup Relations*, 9(1), 25–42. <https://doi.org/10.1177/1368430206059856>
- Diel, A., Weigelt, S., & MacDorman, K. F. (2021). A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Transactions on Human–Robot Interaction*, 11(1), 1–33. <https://doi.org/10.1145/3470742>
- Diemer, J., Alpers, G. W., Peperkorn, H. M., Shibana, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, 6(26), 26. <https://doi.org/10.3389/fpsyg.2015.00026>
- Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *The British Journal of Social Psychology*, 51(4), 724–731. <https://doi.org/10.1111/j.2044-8309.2011.02082.x>
- Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4), 1494–1509. <https://doi.org/10.1016/j.chb.2007.05.007>
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125–130. <https://doi.org/10.1016/j.cognition.2012.06.007>
- Hepperle, D., Ödell, H., & Wölfel, M. (2020). Differences in the uncanny valley between head-mounted displays and monitors. In *2020 International Conference on Cyberworlds (CW)* (pp. 41–48). IEEE.
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9), 641–661. <https://doi.org/10.1016/j.ijhcs.2008.04.004>
- Jentsch, E. (1906/2008). *On the psychology of the uncanny. Uncanny modernity: Cultural theories and modern anxieties*. Palgrave.
- Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, 6(6), 390. <https://doi.org/10.3389/fpsyg.2015.00390>
- Kätsyri, J., Mäkäräinen, M., & Takala, T. (2017). Testing the “uncanny valley” hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli. *International Journal of Human–Computer Studies*, 97(January), 149–161. <https://doi.org/10.1016/j.ijhcs.2016.09.010>
- Kim, K., Rosenthal, M. Z., Zielinski, D., & Brady, R. (2012). Comparison of desktop, head mounted display, and six wall fully immersive systems using a stressful task. In *2012 IEEE Virtual Reality Workshops (VRW)* (pp. 143–144). IEEE.
- Kłosowski, D., Kucharska, A., Kwiecień, A., Łupkowski, P., & Wasielewska, A. (2019). *Uwaga, immersja! co bada kwestionariusz immersji?* [Poster presentation]. Poznańskie Forum Kognitywistyczne, Poznań Poland. Retrieved from <https://osf.io/bm2v5>.
- Liebrecht, C., & Hooijdonk, C. (2019). Creating humanlike chatbots: What chatbot developers could learn from webcare employees in adopting a conversational human voice. In *International Workshop on Chatbot Research and Design* (pp. 51–64). Springer.
- Łupkowski, P., Rybka, M., Dziedzic, D., & Włodarczyk, W. (2019). The background context condition for the uncanny valley hypothesis. *International Journal of Social Robotics*, 11(1), 25–33. <https://doi.org/10.1007/s12369-018-0490-7>
- MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; Increasing category uncertainty does not. *Cognition*, 146(January), 190–205. <https://doi.org/10.1016/j.cognition.2015.09.019>
- MacDorman, K. F., & Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interaction Studies*, 16(2), 141–172. <https://doi.org/10.1075/is.16.2.01mac>
- Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146(January), 22–32. <https://doi.org/10.1016/j.cognition.2015.09.008>
- Milcent, A.-S., Kadri, A., & Richir, S. (2022). Using facial expressiveness of a virtual agent to induce empathy in users. *International Journal of Human–Computer Interaction*, 38(3), 240–252. <https://doi.org/10.1080/10447318.2021.1938387>
- Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33–35.
- Noguchi, K., Gel, Y. R., Brunner, E., & Konietzschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, 50(12), 1–23. <https://doi.org/10.18637/jss.v050.i12>
- Palomäki, J., Kunnari, A., Drosinou, M., Koverola, M., Lehtonen, N., Halonen, J., Repo, M., & Laakasuo, M. (2018). Evaluating the replicability of the uncanny valley effect. *Heliyon*, 4(11), e00939. <https://doi.org/10.1016/j.heliyon.2018.e00939>
- Pan, X., & Hamilton, A. F. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3), 395–417. <https://doi.org/10.1111/bjop.12290>
- Piçarra, N. J. G. (2014). *[Predicting intention to work with social robots]* [PhD thesis]. Universidade do Algarve.
- Piwek, L., McKay, L. S., & Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, 130(3), 271–277. <https://doi.org/10.1016/j.cognition.2013.11.001>
- Pochwatko, G., Giger, J.-C., Rózańska-Walczuk, M., Świdrak, J., Kukielka, K., Możaryn, J., & Piçarra, N. (2015). Polish version of the Negative Attitude Toward Robots Scale (NARS-PL). *Journal of Automation Mobile Robotics and Intelligent Systems*, 9(3), 65–72. [https://doi.org/10.14313/JAMRIS\\_3-2015/25](https://doi.org/10.14313/JAMRIS_3-2015/25)

- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Ratajczyk, D. (2022). Shape of the uncanny valley and emotional attitudes toward robots assessed by an analysis of YouTube comments. *International Journal of Social Robotics*, 14(8), 1787–1803. <https://doi.org/10.1007/s12369-022-00905-x>
- Ratajczyk, D., Jukiewicz, M., & Lupkowski, P. (2019). Evaluation of the uncanny valley hypothesis based on declared emotional response and psychophysiological reaction. *Bio-Algorithms and Med-Systems*, 15(2), 20190008. <https://doi.org/10.1515/bams-2019-0008>
- Robinson, H., MacDonald, B., & Broadbent, E. (2014). The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics*, 6(4), 575–591. <https://doi.org/10.1007/s12369-014-0242-2>
- RStudio Team (2020). *RStudio: Integrated development environment for R*. RStudio, PBC. Retrieved from <http://www.rstudio.com/>
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews. Neuroscience*, 6(4), 332–339. <https://doi.org/10.1038/nrn1651>
- Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10(3), 266–281. <https://doi.org/10.1162/105474601300343603>
- Schwind, V., Wolf, K., & Henze, N. (2018). Avoiding the uncanny valley in virtual character design. *Interactions*, 25(5), 45–49. <https://doi.org/10.1145/3236673>
- 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. <https://doi.org/10.1162/pres.16.4.337>
- Sharkey, A., & Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology*, 14(1), 27–40. <https://doi.org/10.1007/s10676-010-9234-6>
- Shin, M., Song, S. W., & Chock, T. M. (2019). Uncanny valley effects on friendship decisions in virtual social networking service. *Cyberpsychology, Behavior and Social Networking*, 22(11), 700–705. <https://doi.org/10.1089/cyber.2019.0122>
- Siebert, R., Taubert, N., Spadacenta, S., Dicke, P. W., Giese, M. A., & Thier, P. (2020). A naturalistic dynamic monkey head avatar elicits species-typical reactions and overcomes the uncanny valley. *eNeuro*, 7(4), ENEURO.0524-19.2020. <https://doi.org/10.1523/ENEURO.0524-19.2020>
- Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. (2019). Help! is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology*, 15(1), 30–54. <https://doi.org/10.17011/ht/urn.201902201607>
- Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the United States of America*, 106(43), 18362–18366. <https://doi.org/10.1073/pnas.0910063106>
- Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160(March), 43–50. <https://doi.org/10.1016/j.cognition.2016.12.010>
- Stein, J.-P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior*, 95(June), 73–82. <https://doi.org/10.1016/j.chb.2019.01.021>
- Strojny, P., & Strojny, A. (2014). Kwestionariusz immersji-polska adaptacja i empiryczna weryfikacja narzędzia. *Homo Ludens*, 1(6), 171–186. <https://ptbg.org.pl/dl/164/>
- Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior*, 27(2), 741–749. <https://doi.org/10.1016/j.chb.2010.10.018>
- Wasielewska, A., & Łupkowski, P. (2021). Nieoczywiste relacje z technologią. przegląd badań na temat ludzkich postaw wobec robotów. *Człowiek i Społeczeństwo*, 51(51), 165–187. <https://doi.org/10.14746/cis.2021.51.9>
- Wechsler, T. F., Kümpers, F., & Mühlberger, A. (2019). Inferiority or even superiority of virtual reality exposure therapy in phobias?—A systematic review and quantitative meta-analysis on randomized controlled trials specifically comparing the efficacy of virtual reality exposure to gold standard *in vivo* exposure in agoraphobia, specific phobia, and social phobia. *Frontiers in Psychology*, 10(September), 1758. <https://doi.org/10.3389/fpsyg.2019.01758>
- Weidner, F., Hoesch, A., Poeschl, S., & Broll, W. (2017). Comparing VR and non-VR driving simulations: An experimental user study. In *2017 IEEE Virtual Reality (VR)* (pp. 281–282). IEEE.
- Yam, K. C., Bigman, Y., & Gray, K. (2021). Reducing the uncanny valley by dehumanizing humanoid robots. *Computers in Human Behavior*, 125(December), 106945. <https://doi.org/10.1016/j.chb.2021.106945>
- Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, 6(June), 883. <https://doi.org/10.3389/fpsyg.2015.00883>
- Złotowski, J., Yogeewaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, 100(April), 48–54. <https://doi.org/10.1016/j.ijhcs.2016.12.008>

## About the authors

**Dawid Ratajczyk** is a cognitive scientist and a PhD student at the Adam Mickiewicz University in Poznań, Poland. Besides a master's degree in Cognitive Science, he also received a bachelor's degree in Biology. His current research interest is focused on social robotics, human-computer interaction, virtual reality, and language processing.

**Jakub Dakowski** studies Cognitive Sciences at Adam Mickiewicz University in Poznań and Computer Science at WSB University in Poznań. His scientific interests include applications of logic, puzzles and new technologies with an emphasis on virtual reality, metaverse, and virtual avatars.

**Paweł Łupkowski** is an Associate Professor at the Faculty of Psychology and Cognitive Science and Co-founder of the Reasoning Research Group. His research interests are philosophical logic (logic of questions), formal analysis of cognitive processes, Human-Robot Interaction, attitudes toward robots, and conceptual foundations of Artificial Intelligence.

### **A.3 Study III: Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality**

## Highlights

### **Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality**

Dawid Ratajczyk, Monika Żyła, Piotr Jaworski, Paweł Łupkowski

- Virtual reality experiment with psychophysiology and self-reports
- Agent appearance is crucial for humanlikeness and eeriness
- Agent identity had no significant impact on eeriness
- Appearance and mind attribution interact to affect likability
- Violating expectancy of experience cues may increase eeriness

# Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality

Dawid Ratajczyk<sup>a</sup>, Monika Żyła<sup>a</sup>, Piotr Jaworski<sup>a</sup>, Paweł Łupkowski<sup>a</sup>

<sup>a</sup>*Faculty of Psychology and Cognitive Science, Adam Mickiewicz University in Poznań, Szamarzewskiego 89/AB, Poznań, 60-568, Poland*

---

## Abstract

This study aimed to investigate the joint contributions of appearance and type of mind attributed to agents on the uncanny valley (UV) effect. An experiment was conducted using an ecologically valid virtual reality environment, where participants engaged in a conversational game, to improve the validity of the results and gain a more comprehensive understanding of the UV effect. The study used a  $2 \times 2$  experimental design, manipulating both the agent's appearance (robot or human) and the identity behind the character (artificial intelligence or a user's mind). The final sample consisted of 116 participants. Self-reported eeriness and likability, as well as, electrodermal and heart activities were measured. It was found that the appearance of the agent was crucial in determining users' perceptions of human likeness and feelings of eeriness, but the type of mind attributed to the agent did not significantly affect feelings of eeriness. Additionally, the interaction between appearance and type of mind attributed to the agent influenced behavioral realism, which, in turn, affected likability. People perceive other users with more humanlike avatars and artificial intelligence presented as robots as more realistic and, thus, more likable. We discuss the possibility that difficulties with emotion recognition and feelings of eeriness are attributed to the inability to recognize emotion, which we identified as a possible cause for uncanny ratings. Our results suggest that cues related to the ability to experience do not increase feelings of eeriness themselves, but rather a violation of the expectancy regarding these cues increases feelings of eeriness. Our study

---

*Email address:* dawid.ratajczyk@amu.edu.pl (Dawid Ratajczyk)

provides insight into the factors that contribute to human-robot interactions and highlights the importance of appearance in designing effective and likable artificial agents.

*Keywords:*

Uncanny valley, Virtual reality, Emotion recognition, Mind attribution, Agent appearance, Humanlike agents, Eeriness

---

## 1. Introduction

In the wake of the latest technological innovations, including the emergence of highly advanced chatbots like ChatGPT<sup>1</sup> and the development of robots capable of replicating human facial expressions, such as Engineered Arts' Ameca<sup>2</sup>, discussions surrounding the effects of artificial agents on individuals and society are increasing in significance. One of the hypotheses that considers such an impact is the uncanny valley hypothesis (UVH), formulated by Mori (1970). The UVH states that imperfectly humanlike artificial characters may elicit an eerie or disconcerting feeling in observers (Burleigh and Schoenherr, 2015). These artificial characters include robots, computer-generated characters, chatbots, and virtual agents.

Virtual agents, in particular, have been the subject of many recent studies in the field of human-computer interactions (e.g., Stein and Ohler, 2017; Wang et al., 2019; Ciechanowski et al., 2019). These agents are computer-generated characters that are designed to interact with humans in a way that is similar to human-human interaction. They can be used for a variety of applications, such as customer service (Söderlund and Oikarinen, 2021), education (Rickel, 2001), and entertainment (Lipp et al., 2021).

Advancements in UVH-related research are essential for understanding the emotions related to human-robot interactions and human-computer interactions, as well as for improving our knowledge on how to develop humanlike characters that are not disturbing to users. By gaining a better understanding of the UV effect, we can ensure that the development of artificial agents is aligned with the needs and preferences of the individuals who interact with them.

---

<sup>1</sup>For an overview see: <https://www.theguardian.com/technology/2023/mar/15/what-is-gpt-4-and-how-does-it-differ-from-chatgpt> [Accessed: 25.06.2024].

<sup>2</sup>See <https://youtu.be/30HILYjsW7c> [Accessed: 25.06.2024].

The UVH has been studied extensively, with numerous researchers examining the effect and proposing various explanations (for a broader perspective, see the meta-analysis of Diel et al., 2021). Despite this, there remains no consensus among scientists regarding the underlying causes of the effect.

In this study, we aim to investigate how the appearance of agents and the type of mind attributed to them (i.e., whether they are perceived as autonomous yet sentient artificial intelligence (AI) or controlled by a human user) jointly contribute to the UV effect. To achieve this, we conducted an experiment using an ecological virtual reality (VR) environment, which allowed us to measure both self-reported and psychophysiological responses to the agents to improve the validity of the results and gain a more comprehensive understanding of the UVH. Properly planned examinations of robots in VR may be even more adequate than studying real robots in laboratories, due to the presence of fewer uncontrolled variables (as a result of strictly programmed events and increased ecological validity).

### *1.1. UV research*

There has been a growing interest in understanding the concept of “humanlikeness” and how it relates to the UVH. To understand the factors that contribute to the UV effect, researchers have explored various dimensions of humanlikeness, including visual and physical aspects, humanlike behavior, and the attribution of mental capacities (Diel et al., 2021). Additionally, the concept of realism is often considered as an independent variable in these studies. In this context, understanding the visual and mental aspects of the UV effect is essential to advance our knowledge of how humans perceive and interact with artificial entities.

#### *1.1.1. Visual aspects of UV effect*

Studies focusing on the visual and physical similarities of artificial agents to humans form a large class of UV-related studies (see Kätsyri et al., 2015). These involve testing explanations based mainly on perceptual approaches, such as categorical ambiguity, perceptual mismatch, and configural processing. Categorical ambiguity occurs when an artificial agent’s features fall in between those of two categories, such as human and non-human, leading to confusion and discomfort in the observer. Perceptual mismatch refers to the discrepancies between the particular features of the appearance of an artificial agent (e.g., artificial eyes on a realistic face), which can lead to a sense of unease. Configural processing involves the processing of a stimulus as a whole

(considering the spatial relations and the global form), and impairments in this process (e.g., presentation of deviations from facial norms) may elicit uncanny feelings.

Many of the studies that have been conducted using these approaches have produced inconclusive results (e.g., Seyama and Nagayama, 2007; Yamada et al., 2013; Burleigh et al., 2013; MacDorman and Chattopadhyay, 2016; Sasaki et al., 2017; Palomäki et al., 2018). An extensive review of the empirical data on the UVH made by Kätsyri et al. (2015) showed that the perception of characters ranging from artificial to human is indeed categorical. However, they did not find a strong support for the hypothesis that ambiguous categorization induces negative affinity. Instead, the review suggested that perceptual mismatch plays a more significant role in influencing emotions related to the UV effect. A recent study by Diel and MacDorman (2021) tested predictions derived from several UVH explanations based on lower-level visual and cognitive processing. According to their results, the configural processing was the best predictor of UV-related effects, and offered a better explanatory potential than the perceptual mismatch hypothesis.

#### *1.1.2. Involvement of mind attribution*

Several studies have provided support for a connection between the UV effect and the theory of mind (e.g., Hegel et al., 2008; Schein and Gray, 2015). A broader explanation and formulation of this theory have been presented by Gray and Wegner (2012). These authors proposed an explanation based on the agency and experience dimensions of mind perception. Agency is the capacity to do, to plan and exert self-control, while experience is the capacity to feel and to sense (see Gray et al., 2007). Gray and Wegner (2012), in their study, demonstrated that a participant's beliefs in a computer's capability to experience triggered the uncanny effect. Furthermore, Stein and Ohler (2017) showed that not only the imaginary objects but perceived interacting characters in VR may be seen as eerier, depending on the beliefs of the observers. If it was thought that the characters were controlled by AI, the participants assessed the characters as eerier than when they thought the characters were controlled by humans. However, the study did not examine any connections with the visual humanlikeness spectrum. Recently, Banks (2020) used five established theory of mind tests to study the mind attribution to robots. Their results indicated that mentalization for robots and for humans is similar when the anthropomorphic social cues of the characters are similar and interpretable for observers. This study also suggested that a

robot’s morphology influences the mind perception via social cues and may evoke heuristics of mentalization. These results indicate that the appearance of robots may influence the attribution of experience and, because of this, some characters seem to be more eerie. In sum, the current research indicates that the mind attribution of robotic agents and its influence on the elicitation of negative emotions is related to human beliefs. However, it remains unknown how these effects relate to the visual aspects of the UV effect.

### *1.2. Humanlikeness and realism*

One underlying assumption in this field is that humans prefer to interact with machines in the same way that they interact with other people. Therefore, researchers tried to create more humanlike robots and more realistic avatars. The distinctions between realism and humanlikeness are not clear. In some studies, these concepts are considered separate (e.g., Phillips et al., 2018), while others treat them as equivalent (e.g., Cheetham et al., 2011). To differentiate between the two concepts, humanlikeness is typically understood in the literature as the property of possessing humanlike features, whether physical or non-physical. On the other hand, realism refers to the degree of fidelity to an object, typically concerning its physical appearance, and is commonly conveyed through mediums such as visual screens or mental representations (cf. Złotowski et al., 2014; Diel et al., 2021; Miller, 2021).

Previous research has shown that the perception of humanlikeness for artificial agents is not a gradual process, but rather a categorical one (Kätsyri et al., 2015). At the same time, we do not know how much these perceptual categories translate into social categories (in terms of psychological essentialism; Demoulin et al., 2006), or how many such categories can be distinguished. Humanlikeness is a multidimensional variable. For instance, the expressed emotionality of a robot can influence its perceived humanlikeness (Złotowski et al., 2014). However, a simple increase in humanlikeness in artificial agents does not necessarily lead to the UV effect (see, e.g., research results obtained by Mathur and Reichling, 2016, and Ratajczyk, 2022). According to the UVH, only highly humanlike robots should elicit feelings of eeriness. Nonetheless, recent studies have suggested the occurrence of two valleys, one for moderately humanlike robots (i.e., artificial agents with a human body, but reduced facial expressions and a non-humanlike surface) and another for highly humanlike robots (i.e., artificial agents with facial expressions and a humanlike surface, but without a human body; Kim et al., 2022). It is currently unclear whether these two valleys are triggered by the same

underlying causes, or if separate explanations are needed for each valley. One possible explanation for both of these valleys is that a humanoid robot fails to meet the expectations of its users. For instance, if a human-shaped robot is expected to perform tasks in a manner similar to a human but fails to do so, then users may experience frustration and disappointment. Similarly, if a robot is designed to express emotions through facial expressions, but the expressions do not align with the user's expectations, it may lead to feelings of distrust and unease.

### 1.3. *Impact of greater immersion*

Many of the UVH studies applied two methodological approaches: utilizing on-screen stimuli or real robots. Both methods are very distinct and have their disadvantages. On-screen stimuli mostly reduce the UV problem only to the visual aspects of stimuli. However, many suggest that in order to properly examine the UVH, there is a need to test real robots and not just images (e.g., Bartneck et al., 2009). On the other hand, real robot studies are rare and very expensive. In addition, it is a challenge to control all of the variables in a real robot interaction scenario. Utilizing VR in UVH studies seems to be a bridge between virtual character and robot studies, due to the immersive nature of this approach. Participants in a properly designed experiment will be able to be *present* in the environment and feel the spatial proximity of robots, with the advantages of a safe experimental set-up and relatively easy manipulations of the visual aspects of characters. Cheetham et al. (2013) have suggested, that 3D visualization may be a good method for testing the UVH. Ratajczyk et al. (2023) have also indicated that the emotions elicited by humanlike characters in VR do not differ significantly from those elicited by characters viewed on-screen. However the immersion and engagement of participants are greater with VR. Level of immersion may be a moderator in the perception of human-computer interactions (Fox et al., 2015).

### 1.4. *Related work*

Several studies have used similar methodological approaches, mainly to study interactions with chatbots.

For example, Nowak (2004) conducted a study with a similar experimental design to the current work where they examined the influence of avatar look and identity (computer or human) on the level of uncertainty and social judgment. For avatar look manipulation, they used low polygonal character

images (early 3D graphics) of humans, non-humans and a no-image control. Identity was manipulated by telling the participants that they were interacting with a human or a computer. The results showed that less anthropomorphic (non-human) characters were more socially attractive and more credible. Identity did not influence these measures. In addition, neither anthropomorphism nor identity influenced the reported levels of uncertainty.

In a meta-analysis conducted by Fox et al. (2015), it was found that human-controlled agents generally have a stronger social influence than computer-controlled agents. Social influence was defined as a multivariable combination of the affective, attitudinal, physiological, and behavioral responses of participants. The results identified several moderators of this relationship, including immersion, the type of affect measurement (subjective vs. objective), and task type. Although the authors acknowledged the potential influence of other variables such as perceived anthropomorphism or the behavioral realism of the representations, their analysis was constrained by the limited number of studies that incorporated these variables. As such, they emphasized the need for further research to explore the impacts of these variables.

Go and Sundar (2019) conducted a study where participants chatted with an agent online. These researchers investigated the impacts of anthropomorphic visual cues, message interactivity, and identity cues on the participants' evaluations of the agent, attitudes toward the website, and behavioral intentions. When the chat agent was represented with a non-human photo, participants in the human identity condition reported lower perceptions of homophily, less favorable attitudes toward the website, and lower behavioral intentions than those in the chatbot identity condition. On the other hand, when the chat agent displayed a highly anthropomorphic visual cues, the participants in the chatbot identity condition showed lower perceptions of homophily, less favorable attitudes, and lower behavioral intentions than those in the human identity condition. These findings suggest an expectancy violation effect where participants' expectations for humanness or machine-like cues were violated, resulting in lower ratings of homophily, attitudes, and behavioral intentions.

Hsu et al. (2024) investigated the potential use of avatars as conversational partners to promote the disclosure of personal information in telehealth. It was found that participants who were interviewed by a shapeless avatar reported more information of a higher quality about a personal event than did participants who were interviewed by a human-appearing avatar.

The participants also preferred the shapeless avatar when discussing embarrassing topics. Knowledge of the avatar’s identity (whether controlled by AI or a human) did not impact the quantity or quality of the reports. These results suggest that an avatar’s appearance plays a more important role than the identity of the questioner in formal medical conversations.

While prior research has examined how appearance and mind attribution influence responses to artificial agents, few studies have explored the joint effects of these factors in immersive contexts. Moreover, experience and agency have been recognized among the most important concepts in mind perception, yet their connection to the uncanny valley’s visual attributes remains understudied. By investigating both avatar appearance and beliefs about who controls them in a VR environment, this study addresses these gaps and offers a more integrated perspective on uncanny responses.

### *1.5. Hypotheses*

The main aim of this study is to investigate how the appearance of an agent (robot vs. human) and the type of mind attributed to it (an autonomous yet emotionally capable AI vs. a human user) interact with each other in an ecological VR environment. Both self-reports and psychophysiological responses are measured.

#### *1.5.1. Is it more important how characters look or what kind of mind we attribute to them for the UV effect?*

Eeriness and likability seem to be the most important dependent variables in the UV effect. Therefore, we decided to examine whether the appearance of the characters and the attribution of a particular type of mind influence these feelings. Previous studies have suggested that moderately humanlike, robotic characters evoke greater eeriness-related emotions and are less likable (Kim et al., 2022). A study that used the same robotic character as in the current research also reported that this character evokes feelings of eeriness (Ratajczyk et al., 2023). Therefore, we have put forth the following hypotheses:

**Hypothesis 1a** *The robotic appearance of a character evokes stronger eeriness feelings of eeriness than a more human appearance.*

The attribution of an ability to experience to AI (one of the factors related to mind attribution in the model of Gray et al., 2007) seems to evoke

feelings of eeriness in people (Gray and Wegner, 2012; Stein and Ohler, 2017). Therefore, we have proposed the following hypothesis:

**Hypothesis 1b** *The attribution of an AI mind to an agent evokes stronger feelings of eeriness than the attribution of a person’s mind.*

We have also formulated a research question regarding the interaction of the appearance and mind attribution factors:

**Research Question 1c** *Does the appearance of a character evoke different levels of eeriness when different types of minds are attributed to them?*

We also formulated the following hypotheses and research question regarding differences in the evoking of likability (with an inverse direction of reaction):

**Hypothesis 2a** *The human appearance of a character evokes stronger likability than a robotic appearance.*

**Hypothesis 2b** *The attribution of a person’s mind to an agent evokes stronger likability than the attribution of an AI mind.*

**Research Question 2c** *Does the appearance of a character evoke different levels of likability when different types of minds are attributed to them?*

### 1.5.2. Psychophysiological results

The overall body of UV research suffers from the variability in self-reported measures used for dependent variables (see discussion in Ratajczyk, 2022). Although emotional self-reports seem to be one of the best measures of the UV effect, an indirect method may complement these affective measures (Diel et al., 2021). Many studies have indicated that psychophysiological measures, such as electrodermal activity (EDA) and heart rate (HR), may be beneficial for human-robot interaction studies (e.g., Weistroffer et al., 2013; Cheetham et al., 2015; Reuten et al., 2018; Kiilavuori et al., 2021), as they can provide an objective, unbiased measure of emotional and cognitive states (Braithwaite et al., 2013).

Psychophysiological measures may be treated as complementary to self-report questionnaires. Higher EDA and HR can indicate greater arousal, emotional reactions, and stress (Boucsein, 2012; Appelhans and Luecken, 2006). Additionally, Van den Hout et al. (2000) showed that fear-related

words produce increased electrodermal reactions. According to the UVH, uncanny characters elicit strong negative feelings. Therefore, an uncanny character should also elicit greater EDA and HR.

However, the results of Kiilavuori et al. (2021) suggest that confrontation with a human appearance evokes a stronger skin conductance response (SCR; one of the EDA measures) than confrontation with a robot. Therefore we formulated the following research questions:

**Research Question 3a** *Does the robotic appearance of a character evoke different psychophysiological reactions (EDA, HR) compared to a human appearance?*

**Research Question 3b** *Does the attribution of an AI mind evoke different psychophysiological reactions (EDA, HR) compared to the attribution of a human mind?*

*1.5.3. Does the type of attributed mind and the appearance of characters influence perceived humanlikeness and realism?*

Humanlikeness may be considered in terms of the physical, behavioral and mental similarities to humans (Diel et al., 2021). Both humanlikeness and realism are considered independent variables in UV research. Thus, it is justified to ask if appearance, the attribution of mind, or an interaction between these factors influence perceived humanlikeness and realism.

**Research Question 4a** *How do the type of attributed mind and the appearance of characters influence humanlikeness?*

**Research Question 4b** *How do the type of attributed mind and the appearance of characters influence realism?*

*1.5.4. What are the reasons behind the emotional reaction?*

To explain how experimental manipulations influence emotional reports and to better understand the results, we formulated the following research questions regarding the mediating effects of the experimental conditions (appearance of characters and attribution of type of mind) on perceived humanlikeness and realism of characters.

**Research Question 5a** *Are the emotional results mediated by the humanlikeness of characters?*

**Research Question 5b** *Are the emotional results mediated by the realism of characters?*

## 2. Method

### 2.1. Experimental set-up

The experiment was performed using VR with the Wizard-of-Oz set-up. The VR environment was designed to replicate a café setting where participants interacted with a character. Inside the VR café, they engaged in a scripted conversational game of “20 questions”, interacting with either a supposed AI or another participant. In reality, all interactions were controlled by a hidden experimenter (wizard). See Figure 1 for the experimental set-up.



Figure 1: Experimental setup for participants. Participants wore VR headset and were placed in a virtual café environment. Psychophysiological data were recorded using the Shimmer3 GSR device, visible on the participant’s left hand, while the VR controller was held in the right hand. A microphone was positioned in front of the participant to enhance the illusion of real-time voice interaction.

Participants wore HTC VIVE Cosmos VR goggles, which feature high-resolution visuals ( $1440 \times 1700$  pixels per eye), a 110-degree field of view, and a refresh rate of up to 90Hz. The experiment was programmed using the Unity game engine<sup>3</sup>.

---

<sup>3</sup><https://unity.com/> [Accessed: 25.06.2024]

We measured psychophysiological data using the Shimmer3 GSR device<sup>4</sup>. Both EDA and optical pulse (PPG) signals were recorded. The Shimmer3 device was mounted on the participants' wrists, the EDA electrodes were mounted on the index and middle fingers, and the optical pulse sensor was either on the little finger or, in the case of small fingers or poor signal acquisition, on the thumbs.

Self-reported questionnaires were collected using a stationary computer after the VR session.

## *2.2. Game interaction*

The primary interaction within the VR environment involved a scripted conversational game of "20 questions". For standardization reasons, we used this game instead of allowing for unrestricted dialogue. The game was played entirely in Polish. The goal was to guess the opponent's word by asking up to 20 yes/no questions, given the category (which was "animals"). The game started with the virtual character asking the participant questions to guess the participant's word, which was always "a ladybug". After the virtual character guessed correctly, roles were reversed, and the participant asked questions to guess the virtual character's word, which was always "a kangaroo".

The participant or the virtual character would continue asking yes/no questions until the given word was guessed correctly or the limit of 20 questions was reached. The virtual character always guessed the word after 14 questions. The virtual character's responses were controlled by the wizard, who played pre-recorded responses appropriate to the participants' questions. Participants used a controller to interact with the VR environment, including drawing the category (which was always a mock-up and not truly randomized) and advancing the game. Information regarding the game, including instructions, the word to be guessed, and the number of questions remaining, was displayed on a virtual panel on the table in front of the participant. The panel was updating in real-time to reflect the current stage of the game and guide the participant on the next actions, ensuring they were aware of their progress. The wizard was responsible for counting the number of questions asked.

---

<sup>4</sup><https://shimmersensing.com/product/shimmer3-gsr-unit/>[Accessed: 25.06.2024]

### 2.3. Stimuli

There were two different characters used for appearance manipulation—a robot and a human<sup>5</sup>. The robotic character used in this study was selected based on prior research showing that moderately humanlike robots—those with humanoid body but reduced facial features—are particularly prone to eliciting uncanny valley effects (Kim et al., 2022). The robotic character in this study had a neutral and mechanical appearance. This design was intentional, as the goal was to investigate how robotic appearance interacts with mind attribution. The human character was chosen to be highly realistic and neutral in appearance. The selection aimed to provide a convincing human without exaggerated or highly stylized features. See Figure 2 for the appearances of characters in the café environment. The characters were animated in the same way (casual subtle movement while sitting). Additionally, the SALSA LipSync Unity plugin<sup>6</sup> was utilized to (a) provide realistic head movements for both characters toward the participant’s head position and (b) to synchronize the lip movements of the human character with its speech. For a mouth equivalent, a red LED bulb was implemented for the robot character, which flickered during its speech.

Besides the experimental characters, there was a bartender at the far end of the room during the entire experiment to increase the immersiveness of the environment.

The verbal utterances of the characters were recorded in a way that they could be considered as generated by a chatbot or a human (i.e., with weak prosody). In this case, the recordings were actually voiced-over by an actor trained to deliver the lines in a manner that conveyed the desired level of artificiality. In total, the wizard could play 46 utterances. These were divided into 4 groups: introductory (e.g., “can you hear me?”, “hello”), guessing in the game (e.g., “does this animal live on land?”, “is this animal an insect?”), answers to participants in the game (e.g., “I think so”, “yes”, and “no” with intonation variations), and auxiliary (e.g., “can you repeat?”, “you can do it!”).

---

<sup>5</sup>The robot model was downloaded for free from [www.mixamo.com](http://www.mixamo.com). The human model was bought from <https://www.daz3d.com/shop/>. The café model was bought from [www.turbosquid.com](http://www.turbosquid.com). All accessed 25.06.2024. Café environment and similar models were used and tested in previous studies (see Ratajczyk et al., 2019, 2023; Łupkowski and Ratajczyk, 2023).

<sup>6</sup><https://crazyminnowstudio.com/docs/salsa-lip-sync/> [Accessed: 25.06.2024]



Figure 2: The two appearance conditions—the human and robot characters—implemented in the café.

To strengthen the attribution of the mental capability of experience, the agent was expressing utterances indicating that it had feelings (i.e., “I feel it will be difficult”, “I’m glad I managed”).

#### 2.4. Study design

The study used a  $2 \times 2$  between-subjects experimental design with two independent variables (IVs). The first IV was *appearance* (human vs. robot), and the second IV was *identity* (attribution of AI’s vs. attribution of user’s mind).

The dependent variables (DVs) included *eeriness*, measured by self-reported ratings, and *likability*, also measured by self-reported ratings.

Additional variables measured were self-reported *humanlikeness*; *behavior realism* and *appearance realism* (both measured with VR Realism questionnaire); and psychophysiological measures including EDA and HR.

### 2.5. Measures

All questionnaires were administered in Polish. In addition to the questionnaires mentioned below, metrics related to age, gender, and education were also collected.

*Eeriness.* UVH studies have used various terms for the dependent variables (Diel et al., 2021). Recent research has shown that the most appropriate way to measure uncanniness is the eeriness variable (e.g., Chattopadhyay and MacDorman, 2016; Kim et al., 2022; Ratajczyk, 2022).

Following the methodology used in previous studies (Gray and Wegner, 2012; Ratajczyk et al., 2023), the eeriness index consisted of a sum value of four intensity questions related to eeriness (Polish translations of “frightening”, “repulsive”, “strange”, “disturbing”<sup>7</sup>). These terms were chosen to properly capture the idea of eeriness in Polish.

- Rate how scary/repulsive/strange/disturbing the character was that you talked to?

Answers were rated on a scale of 1–4: (1) Not at all, (2) A little, (3) Rather, (4) Very much. Cronbach’s alpha for the eeriness questions indicated an acceptable level of reliability ( $\alpha = 0.78$ ).

*Likability.* Previous research has suggested that negative affect (in the form of eeriness) and positive affect associated with the UV effect should be measured separately because they are conceptually different dimensions (Chattopadhyay and MacDorman, 2016; Kim et al., 2022). For positive affect, likability is one of the more commonly used variables in UVH research (e.g., Kätsyri et al., 2015; Mathur and Reichling, 2016; Kim et al., 2022).

Similar to the eeriness rating, we created a likability index comprising the sum value for four intensity questions related to likability (Polish translations of “friendly”, “likable”, “nice”, and “pleasant”<sup>8</sup>). These terms were chosen to properly capture the idea of likability in Polish.

---

<sup>7</sup>Polish original words: “przeróżający”, “odpychający”, “dziwny”, “niepokojący”.

<sup>8</sup>Polish original words: “przyjazny”, “sympatyczny”, “miły”, “przyjemny”.

- Rate how friendly/likeable/nice/pleasant was the character you talked to?

Answers were rated on a scale of 1–4: (1) Not at all, (2) A little, (3) Rather, (4) Very much. Cronbach’s alpha for the likability questions indicated a good level of reliability ( $\alpha = 0.87$ ).

*Humanlikeness.* The participants were asked to assess the humanlikeness of the characters on a 5-point Likert scale. The same question was used in previous studies (Łupkowski et al., 2019; Ratajczyk et al., 2019, 2023). The results of this question were treated as the humanlikeness index.

- How much does the presented model resemble a human?

Answers were rated on a scale of 1–5: (1) Completely not humanlike, (2) Rather not humanlike, (3) It starts to look humanlike, (4) Rather humanlike, (5) Completely humanlike.

*VR Realism Scale.* The VR Realism Scale (Poeschl and Doering, 2013) in the Polish version (Lipp et al., 2021) was used to assess the perceived realism of the VR experience. This scale consists of 14 items rated on a 5-point Likert scale from -2 to 2, with an overall score ranging from -28 to 28. Four subscales have been previously distinguished: Scene Realism (5 items), Audience Behavior (4 items), Audience Appearance (4 items), and Sound Realism (1 item).

For the purposes of this study, we used two subscales for analysis: Audience Behavior (indicating the fidelity of the gestures, postures, and facial expressions of virtual characters), and Audience Appearance (indicating the fidelity and realism of virtual characters’ characteristics, such as the adequateness of clothing, and the general authenticity of virtual characters).

The overall scale showed good reliability when evaluated for internal consistency using Cronbach’s alpha ( $\alpha = 0.82$ ). The Audience Behavior subscale showed good reliability ( $\alpha = 0.81$ ), while the Audience Appearance subscale showed acceptable reliability ( $\alpha = 0.74$ ).

### 2.5.1. Psychophysiological measures

We analyzed the following EDA measures: mean of the tonic component of EDA, variance of the phasic component of EDA, frequency of spontaneous SCR, and event-related SCR (ER-SCR). The first three measures were analyzed over the entire interaction with the character, whereas ER-SCR was

assessed as reactions to character utterances. EDA results from sympathetic neuronal activity and may be an indicator of emotional and cognitive states (Braithwaite et al., 2013).

PPG was used to measure HR in the form of beats per minute (BPM). HR may be considered as indicator of arousal, task involvement, and anxiety (Jorna, 1992).

### *2.5.2. Manipulation effectiveness check*

At the end of the experiment, a manipulation effectiveness check was conducted by the experimenter to ensure that the deceptive instructions had successfully influenced participants' beliefs about their interaction partners. Participants were asked to state whether they believed they were interacting with a human or an AI during the game. This check helped confirm the effectiveness of the experimental manipulation in creating the intended beliefs about the identity of the interaction partner.

## *2.6. Participants*

The participants were recruited *via* a paid advertisement on a Facebook website, targeting adults from Poznań and nearby areas, where the laboratory is located. The advertisement invited individuals to take part in a scientific study examining conversational capabilities and participant impressions in VR. The participants had to be at least 18 years old, native Polish, and have no experience with our previous studies. Before the formal study was initiated, the procedure was tested on 9 participants in a pilot study. Following this, minor procedural changes were made (i.e., extending deceptive cues).

Overall, 127 participants took part in the study. The data from 1 participant was not saved. Five participants were also excluded due to technical issues or disruptions in the experimental procedure. Additionally, 5 people were excluded after the aforementioned manipulation effectiveness check. The final sample consisted of 116 participants (63 women, 52 men, 1 unspecified), with an average age of 26.9 years ( $SD = 7.9$ ). During the psychophysiological analysis stage, it was necessary to exclude another 4 participants from the EDA analysis and 5 participants from the photoplethysmogram (PPG) analysis due to measurement disruptions.

Each participant received a gift card for an e-commerce platform valued at 50 PLN (approximately 11 USD) for their participation in the study.

### *2.7. Procedure*

The participants first provided informed consent and received information about the procedures and a technical briefing about the VR headset. The real purpose of the study was not revealed to the participants. They were told that “the study examines the possibilities of a new chat environment within VR where you can chat in real-time”. During this introductory phase, participants practiced the “20 questions” game with the experimenter.

The participants were randomly assigned one of two possible appearances for the character—either a robot character or a human character. They received deceptive instructions indicating that they would play a conversational game in a virtual environment with a virtual character—either an avatar controlled by AI or another participant taking part in the experiment in a different room. This resulted in four different conditions. In fact, in all cases the conversation was controlled by the wizard, who was blind to the conditions.

Before beginning the experiment, we asked participants to read a single-page mock-up newspaper article about VR and its applications. The articles are available in the supplementary materials. In the AI-controlled condition, the participants received an article with additional lines of text telling them that “In modern systems, artificial intelligence is used, which can perfectly empathize with the role of a human. These are systems based on artificial neural networks that can conduct dialogue in real time, based on automatic speech processing, emotional algorithms and social cognition. Virtual artificial characters show a surprising level of empathy”. They were also verbally informed that “AI systems could now process dialogue in real-time, recognize emotions in speech, and react the way humans do”. This procedure was inspired by the one used by Stein and Ohler (2017).

To measure psychophysiological data, a Shimmer3 GSR device was mounted on the participants’ wrists, and PPG sensor and EDA electrodes were mounted on the participants’ fingers. After the signal collection was tested, the participants received assistance with putting on the VR headset. If necessary, the headset was adjusted for clear vision.

Next, the participants were asked to choose an avatar character to represent them and were informed that the software changed their voices to make them match the selected avatar. This additional deception was added after the pilot study and aimed to make the pre-recorded utterances sound like believable utterances of a human (whose voice was processed by software), or an artificial agent. The participants were then placed inside a virtual

café in front of their opponent. After exchanging greetings, the participant played a conversational game of “20 questions”. The opponent always started by querying the participant. The order of the questions was always the same and the opponent guessed the word correctly at all times. Subsequently, the roles were reversed and participants had to make their guesses knowing only the category of a word. If the word was guessed correctly, the person won<sup>9</sup>. The game lasted 7 minutes on average ( $SD = 2.4$ ).

The participants were told that the words in the game were randomly selected. However, the game was actually prepared and scripted before the experiment. The participants’ answer was always “a kangaroo” and the opponent’s word to guess was always “a ladybug”. The wizard was choosing previously recorded utterances adequate to the script of the conversation and suitable for the participants’ questions.

In the human-controlled conditions, the participants listened to a brief simulated conversation between the experimenter and a fictional experimenter conducting the study in another laboratory, both before and after the game.

After the VR section, the participants were asked to assess the human-likeness, likability, and eeriness of the characters and to complete the VR Realism Scale (Lipp et al., 2021), the Beliefs in Human Uniqueness Scale (Pochwatko et al., 2015; not analyzed in this article), and “the Big Five” personality questionnaire (Strus et al., 2019; not analyzed in this article). In the end, the participants were briefed about the true purpose and procedures of the experiment. To check the deception manipulation, the experimenter inquired about the participants’ beliefs regarding the condition. Five participants from the user condition were excluded from the analysis because they reported being convinced that they were communicating with a non-human entity (the manipulation effectiveness check). The experimental procedure was approved by the Institutional Ethics Committee at Adam Mickiewicz University in Poznan. Refer to Figure 3 for the study design.

### 3. Results

Initially, two-way ANOVAs were conducted to test the hypotheses concerning the effects of agent appearance (human vs. robot) and the type of

---

<sup>9</sup>Winning the game had no statistical effect on the dependent variables of the study. A Mann Whitney U test revealed no significant difference in likability  $U = 1288.5$ ,  $p = 0.141$ , or eeriness  $U = 1538.0$ ,  $p = 0.993$  between participants who win and those who lost.

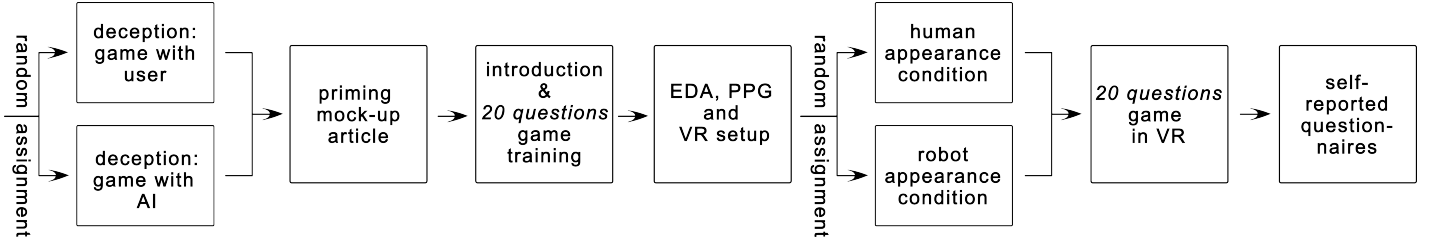


Figure 3: Schematic representation of the study design.

mind attributed (AI vs. human user) on self-reported measures of eeriness and likability. Following this, we investigated the psychophysiological data. The analysis included the use of ANOVA to assess differences in mean tonic EDA, variance of the phasic EDA, frequency of spontaneous SCR, ER-SCR, BPM across conditions. Lastly, mediation analyses were performed to determine whether perceived humanlikeness and realism acted as mediating variables in the relationship between the experimental manipulations and the participants’ emotional responses.

*3.1. Is it more important how characters look or what kind of mind we attribute to them for the UV effect?*

### *3.1.1. Eeriness*

There were 4 independent groups of participants (appearance condition  $\times$  identity condition;  $2 \times 2$ ). Due to non-normal distributions of the data and unequal variances, we used aligned rank transformation (ART) to perform nonparametric factorial data analysis (Wobbrock et al., 2011) and ANOVA procedure. We performed a  $2$  (appearance condition: human and robot)  $\times 2$  (identity condition: AI and user) test with eeriness as the dependent variable. Neither the main effect for the identity condition,  $F(1, 112) = 0.93$ ,  $p = 0.337$ , nor the interaction between conditions,  $F(1, 112) = 1.07$ ,  $p = 0.304$ , were significant. However, there was a significant main effect for the appearance condition,  $F(1, 112) = 3.95$ ,  $p = 0.049$ ,  $\eta_p^2 = 0.034$ , indicating a small effect of this variable on eeriness (see Figure 4).

Thus, hypothesis (1a) was confirmed, as the robotic appearance of the character evoked stronger feelings of eeriness than a human appearance. Hypothesis (1b) was not confirmed; therefore, we cannot say that the attribution of an AI mind evokes stronger feelings of eeriness than the attribution of a human mind. The answer to research question (1c) is negative, as the ap-

pearances of the characters did not evoke different levels of eeriness when different types of minds were attributed to them.

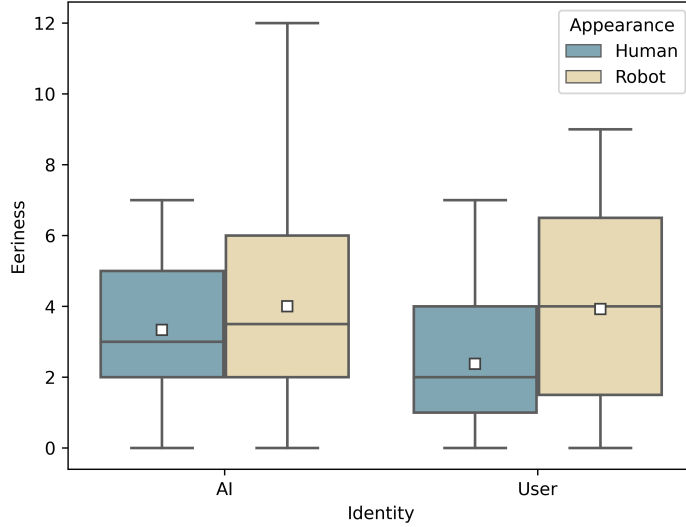


Figure 4: Eeriness levels across the conditions. The lower and upper boundaries of the box represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles, respectively. The line inside the box represents the median. Squares indicate the mean, while the lower and upper error lines depict the minimum and maximum values, respectively.

### 3.1.2. Likability

Due to the non-normal distribution of the data, we used ART to carry out a nonparametric factorial data analysis and ANOVA procedure. We performed a 2 (appearance condition: human and robot)  $\times$  2 (identity condition: AI and user) test with likability as the dependent variable. A two-way ANOVA revealed that there was a statistically significant interaction between the appearance condition and the identity condition ( $F(1, 112) = 6.05$ ,  $p = 0.015$ ,  $\eta_p^2 = 0.051$ ) with a small effect size.

The main effects were not significant for the identity condition,  $F(1, 112) = 1.26$ ,  $p = 0.264$ , or the appearance condition,  $F(1, 112) = 0.80$ ,  $p = 0.373$ , (see Figure 5).

Post hoc Tukey tests revealed no significant differences between particular pairs of conditions.

Thus, hypothesis (2a) was not confirmed, as a human appearance of the character did not evoke greater likability than a robotic appearance (as a sole variable). Hypothesis (2b) was also not confirmed as the attribution of a human mind did not evoke a stronger likability than the attribution of an AI mind. The answer to research question (2c) is positive—the appearances of characters evoke different levels of likability when different types of minds are attributed to them. In particular, robot-looking AIs and human-looking users are more likable than robot-looking users and human-looking AIs.

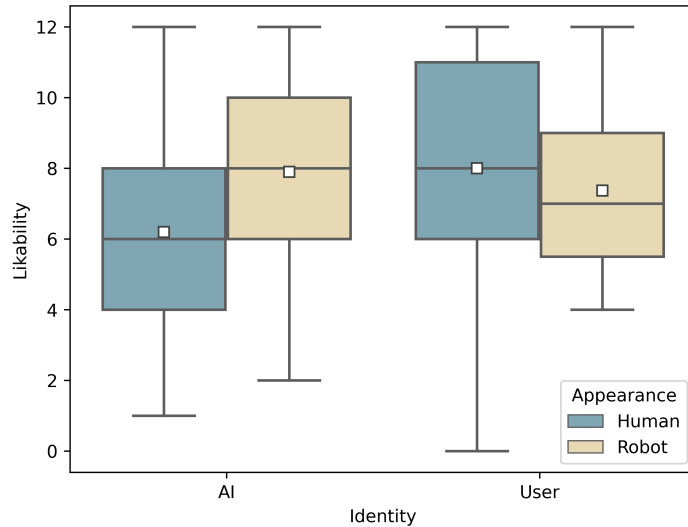


Figure 5: Likability levels across the conditions. The lower and upper boundaries of the box represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles, respectively. The line inside the box represents the median. Squares indicate the mean, while the lower and upper error lines depict the minimum and maximum values, respectively.

## 3.2. Psychophysiological results

### 3.2.1. General EDA

We analyzed EDA starting from the moment the participants were transferred to the café until the last utterance spoken by the character ( $M = 7.0$  minutes,  $SD = 2.4$ ).

The signal was processed and cleaned using the NeuroKit2 Python package (version 0.2.2). We extracted the average tonic EDA, the variance of

Table 1: ANOVA test results for the electrodermal activity (EDA) measures. In all cases, the degrees of freedom = 1,108. SCR = skin conductance response, Var = variance.

	Mean Tonic EDA	Var Phasic EDA	Frequency of Spontaneous SCR
Identity	$F = 0.23, p = 0.63$	$F = 0.33, p = 0.57$	$F = 0.11, p = 0.74$
Appearance	$F = 0.13, p = 0.72$	$F = 2.07, p = 0.15$	$F = 0.29, p = 0.59$
Interaction	$F = 0.66, p = 0.42$	$F = 0.30, p = 0.58$	$F = 0.01, p = 0.92$

the phasic EDA, and the frequency of spontaneous SCRs. The average tonic EDA and variance of the phasic EDA data were not suitable for parametric analysis due to the non-normal distribution of the data. Therefore, we transformed them using ART.

We performed a 2 (appearance condition: human and robot)  $\times$  2 (identity condition: AI and user) ANOVA tests with each measure (i.e., mean tonic EDA, variance of the phasic EDA, frequency of spontaneous SCR) as the dependent variable. None of the results were significant (see Table 1).

### 3.2.2. Reaction to utterances of the characters

We analyzed ER-SCR in response to the verbal utterances of the characters. Segments of the signal were extracted that corresponded to the onsets of the character’s verbal utterances. The ER-SCR signal was baseline-corrected with the minimal value of samples within 900-3500 ms after an utterance onset (see Sjouwerman and Lonsdorf, 2019). Utterances overlapping with previous utterances were removed from the analysis (i.e., if an utterance started within 10 seconds of a previous utterance, then it was removed from the analysis). Additionally, reactions to short utterances from the phase of the experiment where participants were guessing the answer (‘yes’, ‘no’) were removed<sup>10</sup>. On average, 5.56 utterances per participant were analyzed ( $SD = 2.77$ ). The changes in the EDA signal after triggers are presented in Figure 6.

To statistically test for differences in ER-SCRs between the conditions and to examine the interaction, we recorded the maximal value of conductance within 900-6000 ms after the stimulus onset (900 ms is the quickest possible onset of reaction (Sjouwerman and Lonsdorf, 2019), and the reaction

<sup>10</sup>Reactions for such utterances were most likely related to the content of the “20 questions” game.

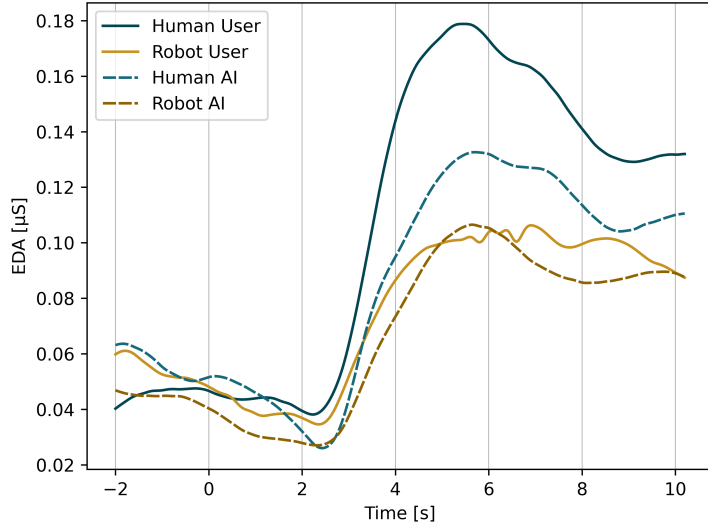


Figure 6: Skin conductance response reactions to the character utterances averaged across conditions. EDA = electrodermal activity.

builds up to 5-6 seconds (Ciechanowski et al., 2019; Bach et al., 2010)).

We conducted an ANOVA on the SCR data that was transformed by ART<sup>11</sup>. A 2 (appearance condition: human and robot)  $\times$  2 (identity condition: AI and user) test with maximal SCR as the dependent variable showed a significant main effect for the appearance condition,  $F(1, 606) = 12.13$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.020$ , with a small effect size, a significant main effect for the identity condition,  $F(1, 606) = 4.18$ ,  $p = 0.041$ ,  $\eta_p^2 = 0.007$ , with a negligible effect size, and a significant interaction between the conditions,  $F(1, 606) = 7.76$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.013$ , with a small effect size.

Post hoc Tukey tests revealed significant difference between robot-looking AI and human-looking user ( $p = 0.008$ ), between robot-looking user and human-looking user ( $p = 0.001$ ), and between human-looking AI and human-looking user ( $p = 0.026$ ). None of the other comparisons were significant.

<sup>11</sup>The residuals were not normally distributed, and there was no homogeneity of variance.

### 3.2.3. Heart activity

PPG data was filtered using a 4-order Butterworth lowpass filter with a 5 Hz cut-off frequency. The data was then processed (including peaks enhancement) using the Heartpy Python package (version 1.2.7). We analyzed the signal starting from the moment the participant was transferred to the café until the last utterance spoken by the character. The analysis was performed on BPMs.

The BPM data were transformed by ART before the ANOVA test due to the non-normal distribution of the data.

A 2 (appearance condition: human and robot)  $\times$  2 (identity condition: AI and user) ANOVA test with BPM as the dependent variable showed a significant main effect for the identity condition,  $F(1, 107) = 7.05$ ,  $p = 0.009$ ,  $\eta_p^2 = 0.062$ , with a moderate effect size (see Figure 7).

Neither the main effect for the appearance condition,  $F(1, 107) = 0.79$ ,  $p = 0.377$ , nor the interaction between conditions,  $F(1, 107) = 0.02$ ,  $p = 0.893$ , were significant.

### 3.2.4. Summary of psychophysiological results

The results show that the answer to research question (3a) is positive only for ER-SCR (i.e., a robotic appearance evokes a smaller SCR than a human appearance, and a human-looking user evokes a greater SCR than a human-looking AI). The answer to research question (3b) is positive only for heart rate (BPM). HR was greater in the user condition.

## 3.3. Does the type of attributed mind and the appearance of characters influence perceived humanlikeness and realism?

### 3.3.1. How do the type of attributed mind and the appearance of characters influence humanlikeness?

We examined if the humanlikeness assessment differed between the experimental groups. Due to the non-normal distribution of the data, we used ART for nonparametric factorial data analysis and ANOVA procedure. We performed a 2 (appearance condition: human and robot)  $\times$  2 (identity condition: AI and user) test with humanlikeness as the dependent variable. A two-way ANOVA revealed that there was a statistically significant main effect for the appearance of the characters,  $F(1, 112) = 58.39$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.34$ , with a strong size effect. The average humanlikeness for the robot character was 2.4 ( $SD = 0.9$ ) and average humanlikeness for human character is

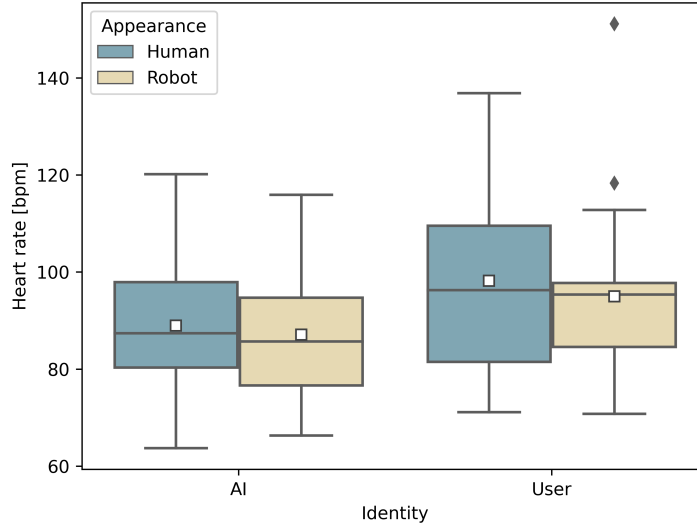


Figure 7: Heart rate levels for conditions. The lower and upper boundaries of the box represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles, respectively. The line inside the box represents the median. Squares indicate the mean, while the lower and upper error lines depict the minimum and maximum values, respectively. Outliers are marked by diamonds. BPM = beats per minute.

3.7 ( $SD = 1.0$ ). See Table 2 for the means and standard deviations for all groups.

The main effects were not significant for the identity condition,  $F(1, 112) = 1.82$ ,  $p = 0.180$ , nor for the interaction,  $F(1, 112) = 2.99$ ,  $p = 0.087$ .

Answering research question (4a), only the appearance of the characters influenced perceived humanlikeness. Characters with a human appearance are perceived to be more humanlike than those with a robot appearance. The attribution of different types of mind did not influence perceived humanlikeness.

### 3.3.2. How do the experimental conditions influence appearance realism?

We also examined if the appearance realism in VR differed between experimental groups. Due to the non-normal distribution of the data and lack of variance homogeneity, we used ART for the nonparametric factorial data analysis and ANOVA procedure. We performed a 2 (appearance condition:

human and robot)  $\times$  2 (identity condition: AI and user) test with appearance realism as the dependent variable. A two-way ANOVA revealed that there was a statistically significant main effect of the appearance condition,  $F(1, 112) = 25.27$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.18$ , with a strong effect. The average appearance realism for the robot character was 1.6 ( $SD = 4.0$ ) and average appearance realism for the human character was 4.6 ( $SD = 2.2$ ). See Table 2 for the means and standard deviations for each group.

The main effect was not significant for the identity condition,  $F(1, 112) = 3.06$ ,  $p = 0.083$ , nor was there a significant interaction,  $F(1, 112) = 2.66$ ,  $p = 0.106$ .

### *3.3.3. How do the experimental conditions influence behavioral realism?*

We examined if behavioral realism in VR differed between the experimental groups. Due to the non-normal distribution of the data and lack of variance homogeneity, we used ART for the nonparametric factorial data analysis and ANOVA procedure. We performed a 2 (appearance condition: human and robot)  $\times$  2 (identity condition: AI and user) test with behavioral realism as the dependent variable. A two-way ANOVA revealed that there was a statistically significant interaction between the appearance condition and the identity condition,  $F(1, 112) = 5.88$ ,  $p = 0.017$ ,  $\eta_p^2 = 0.050$ , with a small effect size. See Table 2 for the means and standard deviations for each group.

The main effects were not significant for the identity condition,  $F(1, 112) = 0.02$ ,  $p = 0.882$ , or the appearance condition,  $F(1, 112) = 0.07$ ,  $p = 0.790$ .

Post hoc Tukey tests showed no significant differences between particular pairs of conditions.

Answering research question (4b), the appearance of the characters influences appearance realism. Human characters are perceived as having a more realistic appearance than robot characters. Attributed identity, together with character appearance, influences behavioral realism. Robotic characters with AI minds and human characters with user minds are perceived as having more realistic behavior than robotic characters with user minds and human characters with AI minds.

Table 2: Means, standard deviations and ranges for behavioral realism, appearance realism, and humanlikeness across conditions.

		Robot user	Robot AI	Human user	Human AI
Behavioral realism	mean	-1.5	0.0	0.1	-1.9
	SD	3.9	4.2	3.3	3.4
	range	[-8,8]	[-8,7]	[-6,7]	[-7,4]
Appearance realism	mean	0.5	2.7	4.6	4.6
	SD	4.4	3.4	2.1	2.3
	range	[-6,8]	[-6,8]	[0,8]	[-1,8]
Humanlikeness	mean	2.2	2.7	3.7	3.6
	SD	1.0	0.8	0.9	1.0
	range	[0,3]	[0,3]	[0,4]	[1,4]

### 3.4. What are the reasons behind the emotional reaction?

#### 3.4.1. Are emotional results mediated by the humanlikeness of the characters?

As there were differences in eeriness (sec. 3.1.1) and humanlikeness (sec. 3.3.1) between the appearance conditions, we examined whether humanlikeness mediates the effect of appearance on eeriness. To examine this, we used the procedure proposed by Baron and Kenny (1986).

The effect of appearance manipulation (human vs. robot) was fully mediated via perceived humanlikeness. As Figure 8 illustrates, the regression coefficients between character appearance and humanlikeness, and between perceived humanlikeness and perceived eeriness, were significant. The indirect effect was  $(-1.24) \times (-0.66) = 0.82$ .

We also tested the significance of this indirect effect using bootstrapping procedures. Unstandardized indirect effects were computed for each of the 10,000 bootstrapped samples, and the 95% confidence interval (CI) was computed by determining the indirect effects at the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles. The bootstrapped unstandardized indirect effect was 0.82 and the 95% CI ranged from 0.16 to 1.49. Thus, the indirect effect was statistically significant ( $p = 0.016$ ).

Thus, the answer to research question (5a) is positive in the case of eeriness. Humanlikeness mediates the influence of character appearance on perceived eeriness.

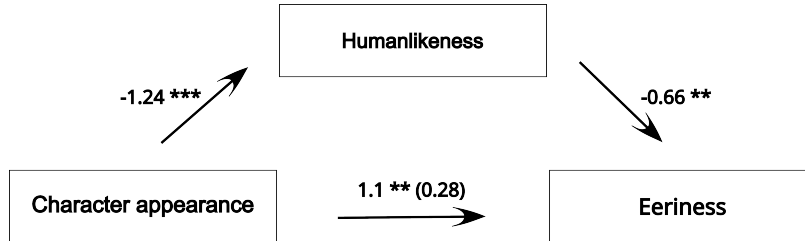


Figure 8: Regression coefficients for the relationship between appearance conditions and eeriness as mediated by humanlikeness. The regression coefficient between appearance and eeriness, controlling for humanlikeness, is in parentheses.  $**p < .01$ ,  $***p < 0.001$ .

#### 3.4.2. Are the emotional results mediated by the appearance realism of the characters?

There were also differences in appearance realism (sec. 3.3.1) between appearance conditions. Therefore, we examined whether appearance realism mediates the effect of appearance on eeriness. The regression coefficient between appearance realism and eeriness was not significant ( $t(2, 113) = -1.2$ ,  $p = 0.23$ ). Therefore, appearance realism does not mediate the effect.

#### 3.4.3. Are the emotional results mediated by the behavioral realism of the characters?

As there were differences in behavioral realism (sec. 3.3.3) and likability (sec. 3.1.2) within an interaction between experimental conditions, we examined if behavioral realism mediates the effect of the interaction between appearance and identity on likability.

To assess whether there is a mediation effect, we conducted a moderated mediation analysis (Muller et al., 2005) using the JSmediation R package, treating appearance as an independent variable and identity as a moderator. This analysis revealed that the identity condition moderates the effect of appearance on behavioral realism,  $t(112) = 2.33$ ,  $p = 0.029$ . It also revealed that the effect of behavioral realism (controlling for the effect of the interaction between appearance and identity) significantly predicted likability,  $t(110) = 4.11$ ,  $p < 0.001$ . This pattern reveals the existence of a moderated mediation. We also computed the first stage moderated mediation index consistent with our joint-significant analysis and it confirmed the moderated mediation, 0.94, 95% CI [0.17, 1.99] (Monte Carlo simulation, 5000 simulations; Yzerbyt et al., 2018; see Figure 9).

Thus, the answer to research question (5b) is positive in the case of be-

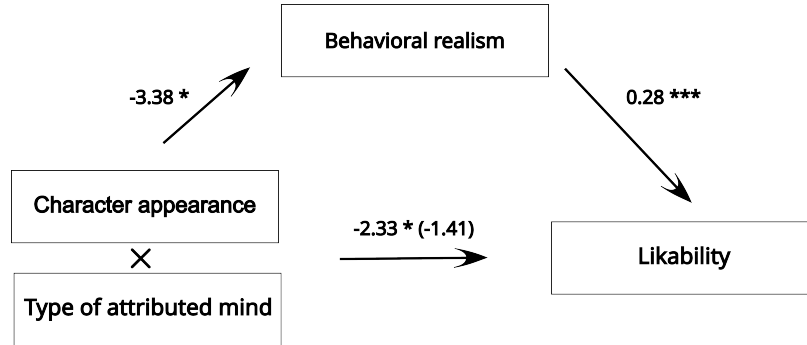


Figure 9: Regression coefficients for the relationship between the interaction of appearance and type of attributed mind and likability as mediated by behavioral realism. The regression coefficient between the interaction and likability, controlling for behavioral realism, is in parentheses.  $*p < 0.05$  ,  $***p < 0.001$ .

havioral realism and likability. Behavioral realism mediates the interaction between character appearance and the type of attributed mind influencing perceived likability.

#### 4. Discussion

The results of this study provide valuable insights into how the appearance and type of mind attributed to agents jointly contribute to the UV effect. The findings suggest that the appearance of an agent is crucial in determining perception of their humanlikeness and feelings of eeriness. However, the type of mind attributed to the agent did not significantly affect feelings of eeriness, indicating that the perception of an agent’s autonomy or control by a human user did not significantly impact the UV effect.

This study also showed that the interaction between the appearance and type of mind attributed to agents influenced behavioral realism, which subsequently affected likability. This suggests that the users with more humanlike avatars and AI presented as a robot are perceived as more realistic and, therefore, more likable. Furthermore, this study revealed that psychophysiological responses, such as SCR and HR, can provide valuable insight into users’ responses to agents in virtual environments.

*4.1. Is it more important how characters look or what type of mind we attribute to them for the UV effect?*

Our study found that the appearance of a character plays a significant role in determining feelings of eeriness. However, the belief that a character represents either AI or a human user did not affect the perceived level of eeriness. Regardless of the attributed type of mind, characters with a robotic appearance were viewed as more eerie than those with a human appearance. Additionally, the interaction between appearance and type of attributed mind influenced likability. Characters with a robot-looking AI and human-looking users were more likable than those with a robot-looking user and a human-looking AI.

Although these results may initially appear conflicting, mediation analysis revealed that appearance influences eeriness through the humanlikeness perception, while the interaction between appearance and type of attributed mind influences likability through behavioral realism. Previous studies have suggested that eeriness and likability are not necessarily opposite measures and may be elicited by different aspects of the human-robot interaction (Zlotowski et al., 2015). Moreover, recent studies have proposed that eeriness may be the best measure for UVH-related emotions (Ratajczyk, 2022; Kim et al., 2022). Based on these findings, it appears that the visual features of characters are more important than the type of attributed mind for the UV effect.

Given the discussions above, it is appropriate to consider emotional self-reports related to eeriness and likability as separate phenomena.

Our results suggest that cues related to the ability to experience do not increase feelings of eeriness. Specifically, priming information about emotional AI, verbal cues given by the experimenter, and emotion-related utterances during the conversation expressed by AI (as well as the user) do not have a significant effect on perceived eeriness. Given the results of Gray and Wegner (2012), cues related to the verification of experience expectations may be primarily captured from visual features, particularly those related to theory of mind, such as facial expressions and gestures. Future research is needed to explore this issue in more detail<sup>12</sup>.

---

<sup>12</sup>The results of the second experiment of Gray and Wegner (2012) can be explained by discrepancies between the mental prototype of artificial agents and the images presented in the story.

#### *4.2. Appearance, humanlikeness and eeriness*

In our experiment, we found that the robot character was perceived as more eerie than the human character. It seems that the robot character falls into the UV for moderately humanlike robots (i.e., artificial agents with a human body but reduced facial expressions and non-humanlike surfaces; see Kim et al., 2022 for a broader explanation). The robot appearance used in our study did not have facial features (only a LED bulb imitating mouth), which can negatively impact an individual’s ability to engage in joint attention and affect other complex aspects of social cognition (Wykowska, 2020). Facial expressions are an important factor in the perception of humanlikeness (Fink, 2012; Phillips et al., 2018). Generally, robots without faces are considered as having a decreased ability to experience, and at the same time, are less preferred to interact with than robots with a displayed humanlike face (Broadbent et al., 2013). Additionally, robots showing empathy and sympathy can improve collaborations without changing the emotional reactions of participants (Winkle et al., 2019). However, previous research has also shown that robots with facial features can be perceived as eerie as well (Gray and Wegner, 2012).

The inability to recognize emotion may be responsible for uncanny ratings (Reuten et al., 2018), and one possible explanation for our results regarding eeriness is the inability to attribute theory of mind to the character. In both experimental conditions (AI and user), the participants had expectations regarding the ability to experience emotions. Human beings are typically expected to have this ability (Gray and Wegner, 2012), and the AI character in our study was described as having emotional capabilities through experimental cues.

Previous findings indicate that individuals tend to have more detailed mental representations of anthropomorphic robots compared to mechanical ones (Kiesler and Goetz, 2002). At the same time, research has shown that when an artificial agent, fails to meet the expectations of participants, it can lead to negative emotions (e.g., Kanda et al., 2004; Go and Sundar, 2019). For instance, if a human-shaped robot is expected to perform tasks in a manner similar to a human but fails to do so, users may experience frustration and disappointment. Similarly, if a robot is anticipated to express emotions through facial expressions but the expressions do not align with the participant’s expectations, it may lead to feelings of unease. Similar results were obtained in the study of Thepsoonthorn et al. (2021) that examined the UV effect from the perspective of a robot’s nonverbal behavior, indicating

that a violation of expected nonverbal behavior, such as gaze, head nodding, and gestures, could induce negative feelings. Also, Schein and Gray (2015) showed that the lack of eyes on a human character was seen as significantly more uncanny than the lack of other facial features and that the effect was partially mediated by perceived emotional experience.

Supporting evidence for this explanation is provided by the EDA results obtained in our study. SCR reactions to character utterances were stronger for agents with a human appearance. Similar results have been obtained by Kiilavuori et al. (2021). Uncanny robots may cause difficulties with emotion recognition along with weaker physiological responses, such as pupil dilation (see Reuten et al., 2018). It has been shown that SCR is stronger in response to faces expressing emotions compared to neutral ones (Banks et al., 2012). Therefore, we suspect that robots with insufficient facial expressions lead to difficulties with emotion recognition, and, furthermore, the feelings of eeriness<sup>13</sup>. The significant differences in ER-SCR observed in the interaction of conditions may be indicative of varying levels of the expected ability to experience, as users are perceived as more sentient than AI despite the use of experimental cues. Therefore, the expectancy violation of exhibiting facial expressions suitable to attributed theory of mind (particularly the ability to experience) may lead to feelings of eeriness.

#### *4.3. Appearance with attributed mind, behavioral realism, and likability*

People perceive robot-looking AI and human-looking user as more likable than human-looking AI and robot-looking user. The robot appearance used in our study did not show any facial expressions. If participants believed that they were seeing such a character representing an artificial mind, they rated it as more realistic in terms of behavior (gestures, postures, and facial expressions) than when they believed the robot was controlled by a user. In contrast, if participants believed that the human character was controlled by an artificial mind, they rated it as having less realistic gestures, postures, and facial expressions, than when they believed the human character was controlled by a user. This means that people expect artificial minds to have a more robotic look and think it is more realistic. It is possible that expectancy

---

<sup>13</sup>Emotion of fear also increases EDA (Van den Hout et al., 2000), however, the eeriness feeling in confrontation with the robot was weak, and we suspect that it was too weak to evoke greater SCR. Moreover, since the suppression of emotions induces stronger SCR (Ohira et al., 2006), the robotic appearance did not lead to emotional suppression.

violations occur in the conditions of human-looking AIs and robot-looking users.

In the absence of disconfirming information, people generally tend to assume that others are similar to themselves (Gordon, 1986). For the user condition, as participants entered the virtual café expecting to meet a user, they might expect to see another person similar to them in appearance. The robotic look violated this expectation. The selection of a representing character is likely to influence how an individual perceives the user, as participants believed that the user had chosen their own character.

For the human-robot interaction research, it is valid to ask why people expect AI to look robotic. It is possible that AI is socially categorized by default as an out-group member (see Eyssel and Kuchenbrandt, 2012). Therefore, an expectation that it will look different from a human will occur. In such a condition, participants expecting to see a non-human character and experiencing one with a human look will violate such an expectation.

#### *4.4. Psychophysiological results, competitiveness and heart activity*

Regarding EDA, there were no significant differences in mean tonic EDA, variance in the phasic EDA, and the frequency of spontaneous SCR between conditions. We hypothesize that the experiment itself was notably stimulating, and the arousal levels of the participants were heightened due to the exposure to new VR technology, potentially masking any differences between conditions. However, a detailed analysis of EDA fragments during the character utterances, when participant’s attention was focused on the character, revealed significant differences between conditions.

HR was found to be higher in the user condition, regardless of character appearance. A meta-analysis conducted by Fox et al. (2015) demonstrated that engaging in competitive or cooperative tasks during experiments with interactions strengthened the effect of agency (user vs. computer) on social influence. When playing competitive games against other humans, individuals may feel a greater emotional investment due to the potential social consequences, such as sharing victory or defeat with others. Similarly, Weibel et al. (2008) showed that participants who played against a human-controlled opponent reported more experiences of presence, flow, and enjoyment in comparison to a computer-controlled opponent. Behnke et al. (2022) showed that emotions did not significantly influence heart activity in gaming scenarios, but being challenged during games led to increased heart activity. Overall,

this suggests that the increased HR in the user condition were caused by greater engagement in the competitive game.

#### *4.5. Limitations*

Although we attempted to recruit a diverse array of participants for our experiment, we found that many of them were highly enthusiastic about new technologies, especially VR. This enthusiasm may have biased their responses to the characters and contributed to the relatively low levels of eeriness observed. In fact, across all groups, the participants tended to view their opponents as more likable than eerie.

While we used a game environment to standardize the experimental procedure, it is possible that the game itself influenced the emotions of the participants. Given that individuals have varying attitudes towards competitive situations, this could have impacted our results.

As highlighted in a meta-analysis by Fox et al. (2015), the identity of the person controlling the character in these experiments can also affect the outcomes. While we attempted to minimize this effect by reducing the wizard's knowledge of the experimental conditions, the fact that the experiment was still conducted by a human may have influenced participants' perceptions of the characters.

Additionally, a limitation of this study is the lack of a pre-experimental baseline physiological data measurement. Due to the deceptive nature of the experiment, participants were exposed to the experimental manipulation as soon as they entered the laboratory, preventing the collection of a resting EDA and HR baseline. As a result, while physiological comparisons within the experimental framework remain valid, absolute values should be interpreted with caution.

A potential limitation of this study is the influence of the specific characters chosen. The effects observed may be related to particular design features rather than generalizable to all humanlike and robotic characters. Future studies incorporating a wider range of robotic and humanlike appearances could help disentangle the effects of specific design choices on emotional perceptions.

To better understand the impact of verbal and non-verbal behavior on the perceptions of robots, future studies should empirically test specific violations of expectations in these domains.

## Acknowledgements

Language models (GPT, Grammarly<sup>14</sup>) were utilized to enhance the writing style and correct language errors. All language correction suggestions were subsequently reviewed and revised by a human.

## Data availability

The Supplementary Materials are available at <https://osf.io/cfmzv/>. The additional datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the National Science Centre, Poland under Grant no. 2020/37/N/HS6/01372.

## References

- Appelhans, B.M., Luecken, L.J., 2006. Heart rate variability as an index of regulated emotional responding. *Review of general psychology* 10, 229–240.
- Bach, D.R., Flandin, G., Friston, K.J., Dolan, R.J., 2010. Modelling event-related skin conductance responses. *International Journal of Psychophysiology* 75, 349–356.
- Banks, J., 2020. Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics* 12, 403–414.
- Banks, S.J., Bellerose, J., Douglas, D., Jones-Gotman, M., 2012. Bilateral skin conductance responses to emotional faces. *Applied psychophysiology and biofeedback* 37, 145–152.

---

<sup>14</sup><https://openai.com/>, <https://grammarly.com/> [Accessed: 25.06.2024].

- Baron, R.M., Kenny, D.A., 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51, 1173.
- Bartneck, C., Kanda, T., Ishiguro, H., Hagita, N., 2009. My robotic doppelgänger—A critical look at the uncanny valley, in: *RO-MAN 2009 – The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 269–276. doi:10.1109/ROMAN.2009.5326351.
- Behnke, M., Gross, J.J., Kaczmarek, L.D., 2022. The role of emotions in esports performance. *Emotion* 22, 1059.
- Boucsein, W., 2012. *Electrodermal activity*. Springer Science & Business Media.
- Braithwaite, J.J., Watson, D.G., Jones, R., Rowe, M., 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1017–1034.
- Broadbent, E., Kumar, V., Li, X., Sollers 3rd, J., Stafford, R.Q., MacDonald, B.A., Wegner, D.M., 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, e72589.
- Burleigh, T.J., Schoenherr, J.R., 2015. A reappraisal of the uncanny valley: categorical perception or frequency-based sensitization? *Frontiers in psychology* 5, 1488.
- 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, 759–771.
- Chattopadhyay, D., MacDorman, K.F., 2016. Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision* 16, 7–7.
- Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., Jancke, L., 2013. Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data. *Frontiers in Psychology* 4, 108. URL: <https://doi.org/10.3389/fpsyg.2013.00108>

[//www.frontiersin.org/article/10.3389/fpsyg.2013.00108](https://www.frontiersin.org/article/10.3389/fpsyg.2013.00108), doi:10.3389/fpsyg.2013.00108.

- Cheetham, M., Suter, P., Jäncke, L., 2011. The human likeness dimension of the “uncanny valley hypothesis”: Behavioral and functional mri findings. *Frontiers in Human Neuroscience* 5, 126. URL: <https://www.frontiersin.org/article/10.3389/fnhum.2011.00126>, doi:10.3389/fnhum.2011.00126.
- Cheetham, M., Wu, L., Pauli, P., Jancke, L., 2015. Arousal, valence, and the uncanny valley: Psychophysiological and self-report findings. *Frontiers in Psychology* 6, 981.
- Ciechanowski, L., Przegalinska, A., Magnuski, M., Gloor, P., 2019. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems* 92, 539–548.
- Demoulin, S., Leyens, J.P., Yzerbyt, V., 2006. Lay theories of essentialism. *Group Processes & Intergroup Relations* 9, 25–42.
- Diel, A., MacDorman, K.F., 2021. Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision* 21, 1–1.
- Diel, A., Weigelt, S., MacDorman, K.F., 2021. A meta-analysis of the uncanny valley’s independent and dependent variables. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 1–33.
- Eyssel, F., Kuchenbrandt, D., 2012. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology* 51, 724–731.
- Fink, J., 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction, in: *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings* 4, Springer. pp. 199–208.
- Fox, J., Ahn, S.J., Janssen, J.H., Yeykelis, L., Segovia, K.Y., Bailenson, J.N., 2015. Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human–Computer Interaction* 30, 401–432.

- Go, E., Sundar, S.S., 2019. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior* 97, 304–316.
- Gordon, R.M., 1986. Folk psychology as simulation. *Mind & language* 1, 158–171.
- Gray, H.M., Gray, K., Wegner, D.M., 2007. Dimensions of mind perception. *science* 315, 619–619.
- Gray, K., Wegner, D.M., 2012. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition* 125, 125–130.
- Hegel, F., Krach, S., Kircher, T., Wrede, B., Sagerer, G., 2008. Theory of Mind (ToM) on robots: a functional neuroimaging study, in: *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, ACM. pp. 335–342.
- Van den Hout, M., de Jong, P., Kindt, M., 2000. Masked fear words produce increased SCRs: An anomaly for Öhman’s theory of pre-attentive processing in anxiety. *Psychophysiology* 37, 283–288.
- Hsu, C.W., Gross, J., Hayne, H., 2024. The avatar face-off: a face (less) avatar facilitates adults’ reports of personal events. *Behaviour & Information Technology* 43, 800–810.
- Jorna, P.G., 1992. Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological psychology* 34, 237–257.
- Kanda, T., Hirano, T., Eaton, D., Ishiguro, H., 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction* 19, 61–84.
- Kätsyri, J., Förger, K., Mäkäräinen, M., Takala, T., 2015. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology* 6, 390.
- Kiesler, S., Goetz, J., 2002. Mental models of robotic assistants, in: *CHI’02 extended abstracts on Human Factors in Computing Systems*, pp. 576–577.

- Kiilavuori, H., Sariola, V., Peltola, M.J., Hietanen, J.K., 2021. Making eye contact with a robot: Psychophysiological responses to eye contact with a human and with a humanoid robot. *Biological psychology* 158, 107989.
- Kim, B., de Visser, E., Phillips, E., 2022. Two uncanny valleys: Re-evaluating the uncanny valley across the full spectrum of real-world human-like robots. *Computers in Human Behavior* 135, 107340.
- Lipp, N., Sterna, R., Dużmańska-Misiarczyk, N., Strojny, A., Poeschl-Guenther, S., Strojny, P., 2021. VR Realism Scale—Revalidation of contemporary VR headsets on a Polish sample. *Plos one* 16, e0261507.
- Łupkowski, P., Ratajczyk, D., 2023. Towards the standardised set of stimuli for the uncanny valley phenomenon studies. *Human Technology* 19, 62–81.
- Łupkowski, P., Rybka, M., Dziedzic, D., Włodarczyk, W., 2019. The background context condition for the uncanny valley hypothesis. *International Journal of Social Robotics* 11, 25–33.
- MacDorman, K.F., Chattopadhyay, D., 2016. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition* 146, 190–205.
- Mathur, M.B., Reichling, D.B., 2016. Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition* 146, 22–32.
- Miller, A., 2021. Realism, in: Zalta, E.N. (Ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2021 ed.. Metaphysics Research Lab, Stanford University.
- Mori, M., 1970. The uncanny valley. *Energy* 7, 33–35.
- Muller, D., Judd, C.M., Yzerbyt, V.Y., 2005. When moderation is mediated and mediation is moderated. *Journal of personality and social psychology* 89, 852.
- Nowak, K.L., 2004. The Influence of Anthropomorphism and Agency on Social Judgment in Virtual Environments. *Journal of Computer-Mediated Communication* 9. doi:10.1111/j.1083-6101.2004.tb00284.x. jCMC925.

- Ohira, H., Nomura, M., Ichikawa, N., Isowa, T., Iidaka, T., Sato, A., Fukuyama, S., Nakajima, T., Yamada, J., 2006. Association of neural and physiological responses during voluntary emotion suppression. *Neuroimage* 29, 721–733.
- Palomäki, J., Kunnari, A., Drosinou, M., Koverola, M., Lehtonen, N., Halonen, J., Repo, M., Laakasuo, M., 2018. Evaluating the replicability of the uncanny valley effect. *Heliyon* 4, e00939.
- Phillips, E., Zhao, X., Ullman, D., Malle, B.F., 2018. What is Human-like? Decomposing Robots’ Human-like Appearance Using the Anthropomorphic roBOT (ABOT) Database, in: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 105–113.
- Pochwatko, G., Giger, J.C., Róžańska-Walczuk, M., Świdrak, J., Kukielka, K., Możaryn, J., Piçarra, N., 2015. Polish version of the negative attitude toward robots scale (NARS-PL). *Journal of Automation Mobile Robotics and Intelligent Systems* 9, 65–72.
- Poeschl, S., Doering, N., 2013. The German VR Simulation Realism Scale—Psychometric Construction for Virtual Reality Applications with Virtual Humans. *Annual Review of Cybertherapy and Telemedicine* 11, 33–37.
- Ratajczyk, D., 2022. Shape of the uncanny valley and emotional attitudes toward robots assessed by an analysis of youtube comments. *International Journal of Social Robotics* 14, 1787–1803.
- Ratajczyk, D., Dakowski, J., Łupkowski, P., 2023. The importance of beliefs in human nature uniqueness for uncanny valley in virtual reality and on-screen. *International Journal of Human–Computer Interaction* 40, 3081–3091.
- Ratajczyk, D., Jukiewicz, M., Lupkowski, P., 2019. Evaluation of the uncanny valley hypothesis based on declared emotional response and psychophysiological reaction. *Bio-Algorithms and Med-Systems* 15, 20190008. doi:<https://doi.org/10.1515/bams-2019-0008>.
- Reuten, A., Van Dam, M., Naber, M., 2018. Pupillary responses to robotic and human emotions: the uncanny valley and media equation confirmed. *Frontiers in psychology* 9, 774.

- Rickel, J., 2001. Intelligent virtual agents for education and training: Opportunities and challenges, in: *Intelligent Virtual Agents: Third International Workshop, IVA 2001 Madrid, Spain, September 10–11, 2001 Proceedings 3*, Springer. pp. 15–22.
- Sasaki, K., Ihaya, K., Yamada, Y., 2017. Avoidance of novelty contributes to the uncanny valley. *Frontiers in psychology* 8, 1792.
- Schein, C., Gray, K., 2015. The eyes are the window to the uncanny valley: Mind perception, autism and missing souls. *Interaction Studies* 16, 173–179.
- 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, 337–351.
- Sjouwerman, R., Lonsdorf, T., 2019. Latency of skin conductance responses across stimulus modalities. *Psychophysiology* 56, e13307.
- Söderlund, M., Oikarinen, E.L., 2021. Service encounters with virtual agents: an examination of perceived humanness as a source of customer satisfaction. *European Journal of Marketing* 55, 94–121.
- Stein, J.P., Ohler, P., 2017. Venturing into the uncanny valley of mind – the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160, 43–50.
- Strus, W., Ciecuch, J., Rowiński, T., 2019. Polska adaptacja kwestionariusza IPIP-BFM-50 do pomiaru pięciu cech osobowości w ujęciu leksykalnym. *Roczniki Psychologiczne* 17, 327–346.
- Thepsoonthorn, C., Ogawa, K.i., Miyake, Y., 2021. The exploration of the uncanny valley from the viewpoint of the robot’s nonverbal behaviour. *International Journal of Social Robotics* 13, 1443–1455.
- Wang, I., Smith, J., Ruiz, J., 2019. Exploring virtual agents for augmented reality, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM. p. 281.

- Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., Groner, R., 2008. Playing online games against computer-vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in human behavior* 24, 2274–2291.
- Weistroffer, V., Paljic, A., Callebert, L., Fuchs, P., 2013. A methodology to assess the acceptability of human-robot collaboration using virtual reality, in: *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, pp. 39–48.
- Winkle, K., Lemaignan, S., Caleb-Solly, P., Leonards, U., Turton, A., Bremner, P., 2019. Effective persuasion strategies for socially assistive robots, in: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE. pp. 277–285.
- Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J., 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures, in: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146.
- Wykowska, A., 2020. Social robots to test flexibility of human social cognition. *International Journal of Social Robotics* 12, 1203–1211.
- Yamada, Y., Kawabe, T., Ihaya, K., 2013. Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research* 55, 20–32.
- Yzerbyt, V., Muller, D., Batailler, C., Judd, C.M., 2018. New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology* 115, 929.
- Złotowski, J., Strasser, E., Bartneck, C., 2014. Dimensions of anthropomorphism: from humanness to humanlikeness, in: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 66–73.
- Zlotowski, J.A., Sumioka, H., Nishio, S., Glas, D.F., Bartneck, C., Ishiguro, H., 2015. Persistence of the uncanny valley: The influence of repeated interactions and a robot’s attitude on its perception. *Frontiers in psychology* 6, 883.

## A.4 Confirmation of authorship statements

Below are five signed statements provided by my co-authors. Each statement relates to one of the publications included in this thesis and confirms the contributions of all authors.

## Co-author declaration: Jakub Dakowski

With reference to Dawid Ratajczyk's application for the **Doctoral degree in Social Communication and Media Studies** at the **Adam Mickiewicz University in Poznań**, I declare that I am a co-author of the publication:

*The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen.*

Authors: Dawid Ratajczyk, Jakub Dakowski, Paweł Łupkowski.

My contribution to this paper includes (10%): Investigation, Resources, Writing – original draft, Writing – review.

*Total contributions by authors:* Dawid Ratajczyk — 80%; Jakub Dakowski — 10%; Paweł Łupkowski — 10%.

Place and date:

Poznań, 4.11.2025

Jakub Dakowski

Signature

(Jakub Dakowski)

## Co-author declaration: Paweł Łupkowski

With reference to Dawid Ratajczyk's application for the **Doctoral degree in Social Communication and Media Studies** at the **Adam Mickiewicz University in Poznań**, I declare that I am a co-author of the publication:

*The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen.*

Authors: Dawid Ratajczyk, Jakub Dakowski, Paweł Łupkowski.

My contribution to this paper includes (10%): Conceptualization, Writing – original draft, Writing – review, Supervision.

*Total contributions by authors:* Dawid Ratajczyk — 80%; Jakub Dakowski — 10%; Paweł Łupkowski — 10%.

Place and date:

Poznań

4.11.25



Signature

(Paweł Łupkowski)

## Co-author declaration: Monika Żyła

With reference to Dawid Ratajczyk's application for the **Doctoral degree in Social Communication and Media Studies** at the **Adam Mickiewicz University in Poznań**, I declare that I am a co-author of the publication:

*Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality.*

Authors: Dawid Ratajczyk, Monika Żyła, Piotr Jaworski, Paweł Łupkowski.

My contribution to this paper includes (10%): Investigation (data collection), Writing – original draft.

*Total contributions by authors:* Dawid Ratajczyk — 75%; Monika Żyła — 10%; Piotr Jaworski — 10%; Paweł Łupkowski — 5%.

Place and date: Poznań, 07.11.2025

Monika Żyła

Signature

(Monika Żyła)

## Co-author declaration: Piotr Jaworski

With reference to Dawid Ratajczyk's application for the **Doctoral degree in Social Communication and Media Studies** at the **Adam Mickiewicz University in Poznań**, I declare that I am a co-author of the publication:

*Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality.*

Authors: Dawid Ratajczyk, Monika Żyła, Piotr Jaworski, Paweł Łupkowski.

My contribution to this paper includes (10%): Investigation (data collection), Resources.

*Total contributions by authors:* Dawid Ratajczyk — 75%; Monika Żyła — 10%; Piotr Jaworski — 10%; Paweł Łupkowski — 5%.

Place and date: Poznań, 12.11.2025y.

Piotr Jaworski

Signature

(Piotr Jaworski)

## Co-author declaration: Paweł Łupkowski

With reference to Dawid Ratajczyk's application for the **Doctoral degree in Social Communication and Media Studies** at the **Adam Mickiewicz University in Poznań**, I declare that I am a co-author of the publication:

*Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality.*

Authors: Dawid Ratajczyk, Monika Żyła, Piotr Jaworski, Paweł Łupkowski.

My contribution to this paper includes (5%): Conceptualization, Writing – original draft, Writing – review, Supervision.

*Total contributions by authors:* Dawid Ratajczyk — 75%; Monika Żyła — 10%; Piotr Jaworski — 10%; Paweł Łupkowski — 5%.

Place and date: Poznań, 4.11.25

Łupkowski

Signature

(Paweł Łupkowski)

## A.5 Biographical Note

### Academic Career

2018–2024	PhD programme in Cognitive Science, Faculty of Psychology and Cognitive Science, Adam Mickiewicz University in Poznań.
2013–2018	M.A. studies in Cognitive Science (M.A. degree), Institute of Psychology, Department of Social Sciences, Adam Mickiewicz University in Poznań.
2011–2014	B.Sc. studies in Biology (B.Sc. degree), Department of Biology, Adam Mickiewicz University in Poznań.

### Scientific Publications

1. Ratajczyk, D., Żyła, M., Jaworski, P., & Łupkowski, P. (in press). Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality. *International Journal of Social Robotics* (accepted October 8, 2025).
2. Ratajczyk, D. J. (2024, May). Dominant or Submissive? Exploring Social Perceptions Across the Human-Robot Spectrum. In *Proceedings of the 2024 4th International Conference on Human-Machine Interaction* (pp. 8–14).
3. Ratajczyk, D., Dakowski, J., & Łupkowski, P. (2023). The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen. *International Journal of Human-Computer Interaction*, 1–11.
4. Łupkowski, P., & Ratajczyk, D. (2023). Towards the standardised set of stimuli for the uncanny valley phenomenon studies. *Human Technology*, 19(1), 62–81.
5. Ratajczyk, D. (2022). Shape of the uncanny valley and emotional attitudes toward robots assessed by an analysis of YouTube comments. *International Journal of Social Robotics*, 14(8), 1787–1803.
6. Łupkowski, P., Danilewicz, O., Ratajczyk, D., & Wasielewska, A. (2022). “It is really interesting how that small robot impacts humans.” The exploratory analysis of human attitudes toward the social robot Vector in Reddit and YouTube comments. *Journal of Automation, Mobile Robotics and Intelligent Systems*, 16(2), 3–10.
7. Jukiewicz, M., Łupkowski, P., Majchrowski, R., Marcinkowska, J., & Ratajczyk, D. (2021). Electrodermal and thermal measurement

- of users' emotional reaction for visual stimuli. *Case Studies in Thermal Engineering*, 27, 101303.
8. Ratajczyk, D. (2019). Uncanny Valley in video games: An overview. *Homo Ludens*, 1(12), 135–148.
  9. Buchwald, M., Kupiński, S., Bykowski, A., Marcinkowska, J., Ratajczyk, D., & Jukiewicz, M. (2019, September). Electrodermal activity as a measure of cognitive load: A methodological approach. In *2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)* (pp. 175–179). IEEE.
  10. Ratajczyk, D., Jukiewicz, M., & Łupkowski, P. (2019). Evaluation of the uncanny valley hypothesis based on declared emotional response and psychophysiological reaction. *Bio-Algorithms and Med-Systems*, 15(2), 20190008.
  11. Ratajczyk, D. (2017). Beginnings of human reasoning: Manifestations of induction, deduction and abduction in hominid activities. *Studia z kognitywistyki i filozofii umysłu*, 10(1).

## Research Grants

- **National Science Centre (Poland), OPUS Grant (2025–2029).**  
*Principal Investigator.*  
Title: “The Effects of Generative Artificial Intelligence on Higher-Order Cognitive Functions”.  
Project no.: 2024/53/B/HS6/02797.  
Funding: 1 662 384 PLN.
- **National Science Centre (Poland), PRELUDIUM Grant (2020–2025).**  
*Principal Investigator.*  
Title: “The uncanny valley in virtual reality – the impact of categorical ambiguity and mind attribution on emotion elicitation”.  
Project no.: 2020/37/N/HS6/01372.  
Funding: 165 078 PLN.