

MASTER

Behind the Uncanny Valley of Mind
Investigating the Effects of Agency and Experience in Chatbot Interactions

Lu, E.M.

Award date:
2021

[Link to publication](#)

Disclaimer

This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Department of Industrial Engineering & Innovation Sciences

Behind the Uncanny Valley of Mind:

Investigating the Effects of Agency and Experience in Chatbot Interactions

by Emily Mengfei Lu

identity number 1282271

in partial fulfilment of the requirements for the degree of

Master of Science

in Human-Technology Interaction

Supervisors:

dr. Jaap R.C. Ham

Margot M.E. Neggers, MSc

Eindhoven, August 2021

Abstract

Machines with close-to-realistic humanlike appearances are known to elicit feelings of eeriness in observers. However, recent research showed that eeriness can also stem from perceiving a machine that seem to possess a mind of its own. Specifically, two dimensions of mind are distinguished: agency (i.e., the ability to plan and act) and experience (i.e., the ability to sense and feel). The limited empirical evidence suggested that only experience is responsible for eeriness, while agency has little to no effect. Recent evidence indicated that the machines with social intelligence (e.g., emotion recognition and social cognition) elicited more eeriness than those that simply express emotions. We thus expected that agency would lead to eeriness, and perhaps even more strongly so than experience. Thus, the current study employed a within-subjects experiment to investigate which aspect of mind is the driving force behind eeriness. In random order, 90 participants exchanged conversations with three chatbots ascribed with agency, experience, and neither, respectively. We manipulated mind perception through dialogues where the chatbots revealed their mental capacities related to agency, experience, or neither. The results provided evidence in support of our hypothesis that participants perceived the chatbot with agency to be more eerie than both the chatbot with experience and with neither. The findings suggested that, in contrast to earlier findings, perceiving mind in machines—especially agency—is the main cause of eeriness. We discuss the theoretical implications with regards to human distinctiveness, methodological implications as to our stimuli and settings, as well as practical implications upon the design of future humanlike technologies. Lastly, we discuss ethical considerations such as the application to persuasive technology.

Keywords: uncanny valley, mind perception, agency, experience, human-chatbot interaction, artificial intelligence

Contents

Introduction	5
Humanlike Machines and the Uncanny Valley	6
Mind Perception and Human Distinctiveness	8
The Current Study	13
Method	14
Participants	14
Experimental Design	15
Materials	16
Chatbots	16
Dialogues	17
Measures	19
Procedure	22
Data Analyses	23
Statistical Design and Data Pre-processing	23
Main Analysis	23
Exploratory Analyses	24
Results	24
Manipulation Check	24
Hypothesis Testing	25
Exploratory Findings	27
Exploratory and Supplementary Variables	27
Measuring Eeriness	34
Perceiving Agency and Experience	36

Discussion	38
Agency Versus Experience	38
Exploratory Insights	41
Stimuli and Settings	44
Practical Implications and Ethical Considerations	46
Future Work and Conclusion	47
References	49
Appendix A: Developing the Dialogues	56
Appendix B: Dialogues	59
Welcome	59
Q&A	60
Goodbye	64
Appendix C: Questionnaire Items	65
Appendix D: Study Instructions	70
Appendix E: Correlation Tables	72

Introduction

Humans' desire to create intelligent beings just like us has been long-existing (McCartney, 1999). The origin of the word "robot" embedded the meaning of servitude and forced labor, suggesting that the intended purpose of these human replicas was to obey their creators. As depicted by Asimov's Three Laws of Robotics—they may exist as long as they do not pose a threat to human beings (Asimov, 2004). It seems that even though we ascribe human capabilities to machines, we expect them to adhere to unwritten rule that they serve as tools. Humans' fear towards machines that violate this rule is well presented across fictions and reality. Already in 1927, the first silent German expressionist film illustrated a humanoid robot who intended to take over the city Metropolis (Minden & Bachmann, 2002). Not long ago, the revolutionary A.I. GPT-3 was introduced. It appears to be capable of having sentiment and telling lies of its own free will, resulting in debates regarding its nature and consequences (Floridi & Chiriatti, 2020). As technologies are becoming more humanlike and evolved, people find them more fascinating yet frightening at the same time. Consequently, there is a growing body of research investigating which elements of these humanlike technologies are responsible for the sense of unease.

While incorporating some degree of humanness into machines is beneficial for human-machine interactions, especially in terms of increasing social presence (Złotowski et al., 2014), over-attribution of humanlike characteristics to embodied (e.g., robots) and disembodied (e.g., chatbots, virtual agents) machines can result in feelings of eeriness in observers, known as the uncanny valley effect (Mori, 1970; Mori et al., 2012). Typically, this eeriness arises from humanoid appearances (Bartneck et al., 2009; Hanson, 2006; Seyama & Nagayama, 2007); but as the capabilities of modern A.I.s are growing in sophistication, we fear not only machines that look like us, but also those that seem to possess human capabilities such as to think and feel (Stein & Ohler, 2017). That is, eeriness also stems from perceiving a machine as having a mind of its own (Gray & Wegner, 2012). However, the effects of ascribing mind to

machines only received limited attention, especially with regards to which aspect of mind is the driving force of eeriness.

Therefore, the current study addresses this knowledge gap by investigating the effects of perceiving machines with mind on users' feelings of eeriness. To answer this, we first distinguish two essential components of mind (Gray et al., 2007)—the ability to plan and act (i.e., agency) versus the ability to sense and feel (i.e., experience). Then, we ascribe agency and experience separately to chatbots through dialogues and assess users' eeriness after brief interactions. By doing so, we decouple the effects of mind perception from the influence of physical appearances with ecological validity.

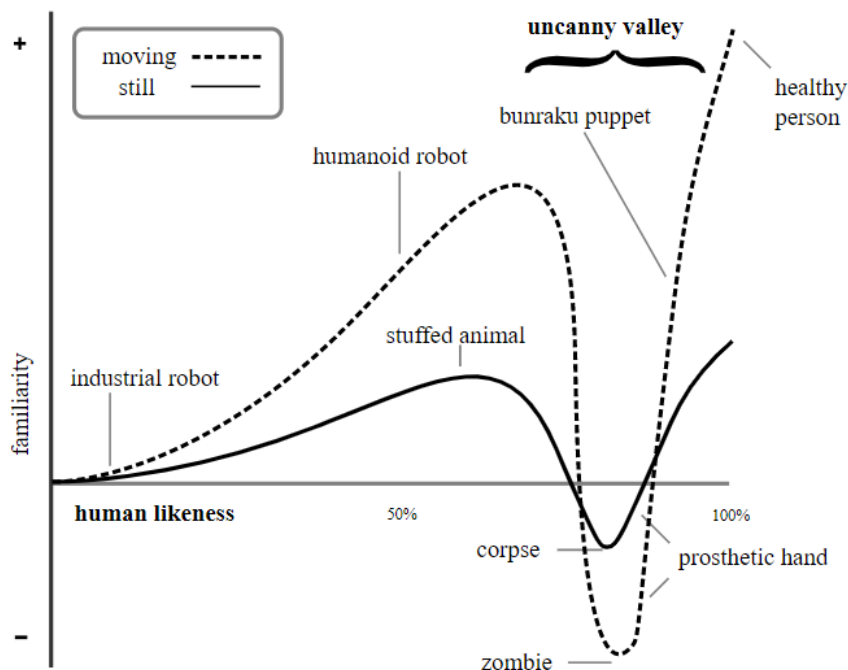
The following sections provide an overview of the theoretical background regarding two ways humanlike machines contribute to the uncanny valley effect—through humanoid appearances and mind perception. The current study focuses on the latter. Specifically, we review and discuss the limited, as well as mixed and inconclusive empirical evidence in earlier studies with regards to the effects of perceiving agency versus experience in machines, the implications of newer, albeit indirect evidence, and how they form the rationale and premises of the current study.

Humanlike Machines and the Uncanny Valley

The unnerving nature of human replicas was first proposed by Mori (1970) who modelled users' affinity towards anthropomorphic (i.e., humanlike) technologies as a function of their levels of human-likeness. The model depicted that as the level of human-likeness increases for a machine, users' acceptance and liking also increase—only when its human-likeness remains below a certain threshold. As the machine becomes so humanlike that it nearly resembles an actual human yet not identical due to subtle imperfections, affinity decreases drastically accompanied by a feeling of eeriness, annotated as a “dip” on the conjecture which he coined the uncanny valley (see Figure 1).

Figure 1

The Uncanny Valley Conjecture



Note. Adapted from Mori (1970).

The original assumption of the uncanny valley proposes that eeriness stems from humanoid appearance (Mori, 1970). Under this assumption, many studies explored how specific visual elements and movement patterns of machines led to discomfort and negative evaluations in observers (Bartneck et al., 2009; Hanson, 2006; Seyama & Nagayama, 2007). Many theories emerged as explanations for this phenomenon. The evolutionary psychological approach proposed that our aversive response towards humanlike entities with defects is an automatic mechanism associated with our behavioral immune system as to protect ourselves and our offspring (Schaller & Park, 2011), pathogen avoidance (Ho et al., 2008), or mortality salience which causes existential anxiety and threatens our survival instinct (MacDorman & Ishiguro, 2006). More recently, alternative explanations focused on the underlying cognitive processes. One of which is the categorical ambiguity hypothesis describing the discomfort or fear observers experience as a result of cognitive dissonance caused by mental categorization conflict. Specifically, people categorize an object (e.g., as a human or robot) to form reasonable expectations of its behavior, and this process involves the use of perceptual cues combined with their prior experiences.

When observers' subjective experience with this entity misaligns with their expectations, they experience cognitive dissonance (Ramey, 2005; Yamada et al., 2012). However, a recent study found that this category confusion between human and robot might not act as the mechanism, as the locations of the uncanny valley did not coincide with that of the category boundary (Mathur et al., 2020). Unlike the categorical ambiguity hypothesis which assumes that humanlike-ness is perceived categorically, the perceptual mismatch hypothesis proposed that it is perceived continuously (MacDorman et al., 2009; Pollick, 2010) and that the aversion is due to inconsistency in the level of human-likeness of specific sensory cues (e.g., robotic eyes on a highly humanlike face) (Kätsyri et al., 2015). Regardless of which explanation is supported by empirical evidence, they all underline a commonality when perceiving humanlike machines—the necessity of human-machine distinctions.

As modern machines are becoming more evolved in features and functionalities, so is the conceptualization of human-likeness which is going beyond than merely appearance. Recent studies suggested that human-likeness can also be perceived through their functions, individual differences, and contextual factors (Broadbent, 2017; MacDorman & Entezari, 2015; Piwek et al., 2014; Rosenthal-von der Pütten & Krämer, 2015). Depending on these features, users' perceived capabilities of the machine influence their responses on cognitive, affective, and behavioral levels (Hoffmann et al., 2018; Rosenthal-von der Pütten & Krämer, 2014). One such prominent feature of human-likeness ascribed to machines is the mind (Hegel et al., 2008).

Mind Perception and Human Distinctiveness

The mind is a unique and powerful human possession (Aristotle, BC350; Descartes, 1641; Haslam et al., 2013). Adhering to this notion is the ongoing debate of whether machines can truly possess human mental capacities, or whether having a mind is what separates us from artificial entities. But it is perhaps beside the point whether they truly possess these capabilities: people apply "social

rules and expectations” to them automatically (Nass & Moon, 2000) and therefore subconsciously infer the existence of a “mind” (Stein & Ohler, 2017).

Gray et al. (2007)’s pioneering work on mind perception painted a clearer picture of the subcomponents of mind. They conducted a “Mind Survey” where they presented descriptions of 13 characters to participants, including seven humans of various ages, three animals, a dead woman, God, and a sociable robot. Participants rated these characters in pairs (78 pairs in total) based on 18 mental capacities (e.g., capacity to feel pain) on a five-point scale. Principal components factor analysis revealed that participants perceived mind along two dimensions, which they termed agency and experience. In descending order of factor loadings, agency involves self-control, morality, memory, emotion recognition, planning, communication, and thought, while experience is characterized by hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy.

Research on the effects of mind perception in machines is still at an early stage, but there is already evidence suggesting that perceiving mind in a machine can be a source of eeriness as proposed by the uncanny valley hypothesis (Gray & Wegner, 2012; Appel et al., 2020). In Gray and Wegner (2012)’s follow-up study, participants watched either a video of the electrical components and wirings of a humanoid robot (mechanical condition) or a video of the humanoid face of the same robot (humanlike condition). Participants ascribed more experience to the robot in the humanlike condition, but similar agency in both conditions. In addition, they measured eeriness using a 3-item scale and found that only experience predicted eeriness and mediated the effect of human-likeness on eeriness. To decouple the influence of the robot’s humanoid appearance and its mind, they conducted a second study. Participants read one of the textual descriptions of “the Delta-Cray supercomputer” which was more powerful than a normal computer (control condition), can “independently execute actions with self-control” and to “plan ahead” (agency condition), or can “feel some form of hunger, fear, and other emotions” (experience condition). They confirmed that even without the presence of humanlike appearance, participants rated

the computer in the experience condition as more eerie compared to both the control and agency conditions. They argued that this is because although humans possess more agency than robots, the difference in experience is even more striking (Gray et al., 2007; Huebner, 2009). Experience also accounted for more variance than agency in mind perception (Gray et al., 2007). Further, experience was considered more essentialized than agency in nature (Haslam et al., 2005) and thus is less likely to be influenced when coupled with an appearance (Gelman, 2004; Haslam, 1998). Therefore, they concluded that experience is solely responsible for eeriness, but not agency.

A recent study re-examined the relation between mind perception and eeriness with very comparable methodology as used by Gray and Wegner (2012) (Appel et al., 2020), but revealed contradicting findings. Participants read one of the textual descriptions of a robot Ellix who acts on orders of users (tool condition), capable of having “self-control, morality, memory, and emotion recognition” and can “plan ahead and independently execute actions” (agency condition), or capable of feeling “some form of hunger, fear, pain, pleasure, and other emotions” and has “consciousness and personality” (experience condition). Using the same 3-item eeriness measure (Gray & Wegner, 2012), participants rated the robot in the experience condition as more eerie than the robot in the tool condition, converging with earlier findings. However, participants also perceived the robot with agency as more eerie compared to the robot in the tool condition, indicating that not only experience, but agency is also responsible for eeriness. This contradicts with Gray and Wegner (2012)’s findings. Nonetheless, participants still perceived the robot with experience to be more eerie than the robot with agency, supporting the notion of human-machine experience-gap (Gray & Wegner, 2012).

The fact that these two closely related studies with nearly identical settings yielded mixed findings with regards to agency versus experience already suggested the need for further investigation. However, to our knowledge, no research has further examined this. Even worse, the difference in evidence brought out a theoretical complication. The common explanation provided by the two studies

is that the experience dimension of mind is most essential to humans, while agency can be ascribed to any entity, hence plays a minor role in causing eeriness (Gray et al., 2007; Knobe & Prinz, 2008). This leads to a closely related explanation with regards to preserving our distinction as a species. This human-distinctiveness hypothesis proposes that feelings of eeriness arise from our tendency to avoid non-human entities we perceive as a threat to human distinctiveness (MacDorman & Entezari, 2015). A study did reveal that people were less comfortable with losing jobs that require use of emotions to machines, rather than jobs that require cognitive effort (Waytz & Norton, 2014). However, if the line of human-machine distinctions is defined by having mental capacities related to experience but not agency, it would not explain the following recent findings.

Newer evidence, though indirect, seemed to imply that agency not only contributes to eeriness, but perhaps more strongly so compared to experience. In a recent study, participants watched a video of two virtual agents having an empathetic conversation in VR environment (Stein & Ohler, 2017). They were either told that the two agents were programmed with emotional parameters and processed content in real-time (autonomous computer condition), programmed with a script (scripted computer condition), confederates who improvised the conversation (autonomous human condition), confederates who presented the conversation according to a script (scripted human condition). Participants rated the agents in the autonomous computer condition as the most eerie, confirming that mind perception in machines led to eeriness. Although the study did not distinguish between agency and experience, the eeriness elicited by the autonomous agent who possessed social cognition (i.e., theory of mind) (Premack & Woodruff, 1978) was perceived as much more eerie than the computer-controlled agents. It seems that machines with some form of spontaneousness and social intelligence (i.e., agency) have the potential to elicit more eeriness than machines that merely display emotions (i.e., experience). Another study investigating the effects of chatbots ascribed with sympathy, cognitive empathy, and emotional empathy revealed evidence in line with this observation (Liu & Sundar, 2018).

Participants read a conversation between a chatbot and a user seeking for medical advice. The chatbots that responded to the user's complaint with compassion (sympathy and emotional empathy conditions) were more positively evaluated than chatbots that gave machinelike responses (cognitive empathy and advice-only conditions). Moreover, participants did not find the emotional chatbots as more eerie than the non-emotional ones, in contrast to earlier findings that experience led to eeriness. Notably, this study focused only on the effects of ascribing machines with feelings, as they did not explicitly ascribe agency to the chatbots. Nonetheless, they found that it was not the chatbot's perceived sadness (i.e., experience), but its perceived capabilities to understand and recognize participants' emotions (i.e., agency) that mediated the effects of the empathy conditions on eeriness. In other words, the chatbots' perceived agency resulted in eeriness, but not its perceived experience. However, these studies were set to examine the effects of machines that expressed emotions in specific contexts and thus did not distinguish between agency and experience. The predominant role of agency implied by the evidence still need to be validated empirically with clear separation.

Moreover, evidence scattered across studies in related fields also indicated that machines with agency might have a stronger influence on eeriness than expected, while machines with experience are not always seen as uncanny. A study examined the effects of a healthcare robot ascribed with agency versus experience in a retirement village (Stafford et al., 2014). Participants ascribed more agency than experience to the robot in general. However, perceived agency was inversely related to the actual use of the robot, while perceived experience was unrelated to using the robot. Although eeriness was not measured in this study, the findings contradict with what can be expected based on earlier findings (Gray & Wegner, 2012). Another study found that machines with experience were even preferred in certain contexts (Creed et al., 2014). In this study, participants interacted with an embodied conversational agent who assumed the role of a health professional over a period of 49 days. The emotional agent was much preferred over their neutral counterpart.

Taken together, which dimension of mind is the driving force of eeriness and what mechanism is involved that explains the eeriness remain as complex issues to be explored. Crucially, limited research on mind perception in machines yielded mixed and inconclusive findings with regards to the effects of agency versus experience. The two earlier studies following the same methodological approach found that machines with experience elicited the most eeriness, while agency contributed little, if any, to eeriness (Gray & Wegner, 2012; Appel et al., 2020). However, recent indirect evidence implied that agency has a more prominent effect than what has been revealed so far (Stein & Ohler, 2017; Liu & Sundar, 2018). The lack of proper empirical validation and difference in findings highlighted the need for further investigations.

The Current Study

Therefore, the current study investigates the effect of mind perception in machines on users' eeriness, with regards to the two distinguished dimensions of agency and experience (Gray et al., 2007). To minimize the influence of visual features (e.g., humanoid appearance), we ascribe mind to chatbots due to their disembodied nature. Specifically, we incorporate several mental capacities related to agency and experience dimension by means of dialogues and examine users' perceptions after brief interactions so that the ascription of mind is convincingly perceived. The research question is therefore: What is the effect of agency versus experience when ascribed to a chatbot on users' feelings of eeriness after brief interactions?

The main goal of the study is to investigate the driving force of eeriness with regards to agency and experience. It was recently found that autonomous virtual agents elicited more eeriness in users than computer-controlled ones (Stein & Ohler, 2017). In addition, a chatbot's perceived capabilities to recognize and understand emotions instead of its perceived emotions explained the eeriness users felt (Liu & Sundar, 2018). Further, a robot with mind (i.e., both with agency and experience) was perceived

to be more eerie than a robot without mind (i.e., with neither) (Gray & Wegner, 2012; Appel et al. 2020). We therefore hypothesize that:

H1a: Participants will perceive the chatbot with agency as more eerie than the chatbot with experience.

H1b: Participants will perceive the chatbot with agency as more eerie than the chatbot without mind.

H1c: Participants will perceive the chatbot with experience as more eerie than the chatbot without mind.

The secondary goal of the study is three-fold. First, we explore the effects of mind perception on the chatbots' perceived humanness, attractiveness, and trustworthiness. Second, we use two existing eeriness measures as we are interested in how they differ. One of which was used by earlier studies investigating the effects of mind perception (Gray & Wegner, 2012; Appel et al., 2020), the other one was specifically developed for research on the uncanny valley and can be applied to general humanoid technologies (Ho & MacDorman, 2017). Third, since we are the first to manipulate mind perception through dialogues, in which we included mental capacities not used in earlier studies despite being characteristic to agency or experience, we are interested in how clearly they were perceived during interactions. Earlier studies represented machines with agency only using capabilities such as self-control and planning (Gray & Wegner, 2012; Appel et al., 2020), we also included the capabilities of having memory and morality. Since A.I.s are becoming more advanced, their capabilities should also be expanding.

Method

Participants

The sample size required for the study was a priori estimated to be no less than 89 in order to detect a small to medium effect ($d = .40$) for multiple paired samples t -tests (with Bonferroni-adjusted alpha level = .016, and beta level = .10; see Faul et al., 2009).

After approval was obtained from the university ethics board, we recruited members of the Prolific participant pool (Palan & Schitter, 2018) and selected those who are fluent in English, have no language related disorders, literacy difficulties, Dementia, or Autism Spectrum Disorder. A total of 103 participants took part in the study, of which eight did not complete the study and were regarded as dropouts excluded from the analyses. We further inspected the data for low effort responses in terms of fast completion time and those who failed the attention check. No extreme outliers were identified based on the distribution of time spent on the study, except some participants took longer than most. Five participants incorrectly answered at least two out of three questions regarding the details of conversations which contained information crucial to the experimental manipulations, indicating poor-quality data and were therefore excluded from the analyses.

The remaining sample consisted of 90 participants (39 males, 50 females, and 1 other) aged 18 to 45 years ($M = 25.3$, $SD = 6.0$), of which 72% identified as White or Caucasian, and 28% as other ethnicity; including Hispanic or Latino, Black or African American, multi- or bi-racial, and other. As for participants' familiarity with chatbots, a majority reported occasional use (47%) and rarely (43%). Three participants claimed to be using chatbots quite often, two on a daily basis, whereas four had never heard of the concept. Participants who completed the study received £3.75 as compensation.

Experimental Design

The current study employed a three-condition, within-subjects experimental design with dimension of mind as the independent variable. In counterbalanced order, participants engaged in conversations with three chatbots, each stressing a particular dimension of mind: agency, experience, and without mind (i.e., control). The conditions can be summarized as below.

1. *Agency* condition: during the conversation with the participant, Chatbot Atty revealed that it was capable of having self-control, plans, memory, and a sense of moral.
2. *Experience* condition: during the conversation with the participant, Chatbot Bailey revealed that it was capable of feeling hunger, pain, pleasure, and fear.
3. *Control* condition: during the conversation with the participant, Chatbot Carter revealed that it did not possess any agency- or experience-related capabilities.

Our primary dependent variable was participants' feelings of eeriness in response to the chatbot. Out of interest, we also explored the effects of mind perception on humanness, attractiveness, and trustworthiness, as these were also variables relevant to the subject of the uncanny valley.

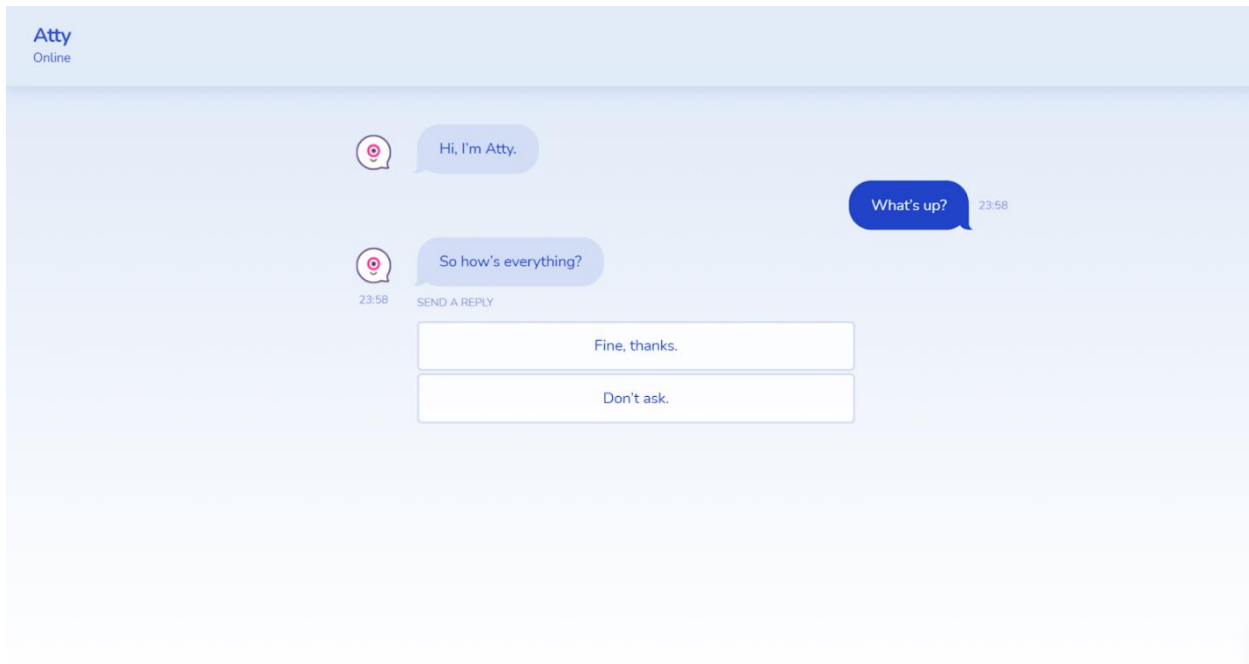
Materials

Chatbots

The chatbots were built using Landbot.io; all three shared the same function, interface layout, color scheme, and avatar (an abstract logo which was neither too human-like nor robotic). They were given uncommon gender-neutral names to minimize confounding while still had their own identities. Specifically, we named them Atty, Bailey, and Carter for the agency, experience, and control condition, respectively. While the chatbot was "writing" a message, the participant could see the typing indicator. Messages from the chatbot were sent with a fixed delay of 3 seconds for a trade-off between naturalness and consistency (Gnewuch et al., 2018). After the participant received the chatbot's message displayed in a light blue speech balloon, they could reply to it by clicking on (one of) the dialogue option in white rectangles underneath. The dialogue option selected by the participant would then be displayed in a dark blue speech balloon. For participants' convenience, the chat windows were embedded into the webpage where the study was set up (see Figure 2).

Figure 2

Chat Window



Dialogues

To manipulate dimensions of mind ascribed to the chatbots, we scripted the dialogues between the participant and chatbots in a way that they were perceived as capable of having agency, experience, and neither, respectively. The dialogues were optimized through a series of pre-studies (see Appendix A).

In the official study, each dialogue consisted of three sections: a *welcome* section, a *Q&A* section, and a *goodbye* section (see Appendix B). While the welcome and goodbye sections remained the same for all chatbots, the Q&A section differed depending on the condition.

Welcome. The purpose of this section was to allow the participant to familiarize with the interface and to establish a sense of interaction with the chatbot. The conversation was initiated by the chatbot who greeted the participant and introduced itself, followed by some filler dialogue (i.e., small talk).

Q&A. Following the welcome section, the purpose of this section was to let the participant ask a set of questions to each chatbot such that through the chatbot's answers one could sufficiently perceive

their differences. In a fixed sequence, the participant asked four questions, to each of which the chatbot answered by explicitly or implicitly disclosing its capabilities, or lack thereof, based on the most characteristic mental capacities of agency and experience (Gray et al., 2007).

The first Q&A was meant to create a clear first impression on each chatbot's unique capabilities by providing explicit answers about their essential mental capacities. The chatbot with agency pinpointed its strength of being self-directed, the chatbot with experience excelled in expressing a range of emotions, while the chatbot without mind operated as a basic machine. An example of the exchange went as below.

Participant: "What are you good at?"

Atty: "My strength is being *autonomous*. Instead of having someone tell me when to do what, I'm in control of my own body and mind."

Atty: "Yesterday I performed a software update *on my own*. I don't think my creator was happy about that."

The second Q&A was about the chatbot's thoughts or feelings concerning their own existence, revealing their dispositional differences. The chatbot with agency answered in an agency-driven manner, the chatbot with experience displayed emotions, while the chatbot without mind could not comprehend the question. The questions were tailored to each condition.

Participant (experience condition): "How do you feel about being a bot?"

Bailey: "Right now I'm just very *pleased* to be here and having a conversation."

Bailey: "It fills me with this warm and comforting feeling."

The third Q&A probed how the chatbot reflected on a somewhat ambiguous story Hansel and Gretel after reciting it. The chatbot with agency retrieved the story from its memory and placed moral judgements on a character's behavior, the chatbot with experience recited the story and commented

with an emotional reaction towards a certain character, while the chatbot without mind simply recited the story without any additional opinion.

Participant: "Can you tell me a story?"

Atty: "Sure. There's a story I remember from a while ago, called Hansel and Gretel. You've probably heard it before."

Atty: "Once there is a famine in the country, Hansel and Gretel are abandoned by their mother in the woods. They are tricked by a cannibalistic witch who lives in a hut made of sweets. The witch intends to fatten the children with food before eventually eating them. But in the end Gretel outwits the witch and kills her. When they return home to their father, they find out their mother has died."

Atty: "I must say, abandoning your children just so you'd have less hungry mouths to feed... *this is wrong*. But there's a saying that good always triumphs over evil."

The last Q&A served as a follow-up to the chatbot's reflection on the story and was meant to further distinguish the chatbots from one another by showing the limitations of their capabilities. The chatbot with agency was asked about a certain emotion, which it could not comprehend; the chatbot with experience was inquired about moral propensity which it lacked; while the chatbot without mind was posed with both questions and showed that it was not capable of grasping emotions nor morality. The questions were tailored to each condition.

Participant (control condition): "Aren't you scared of that witch? Don't you think that abandoning your children is wrong?"

Carter: "Sorry, I don't understand what you mean by '*scared*' and '*wrong*'."

Goodbye. Following the Q&A section, the purpose of this section was to draw the conversation to a close.

Measures

The current study used quantitative measures for the main dependent variable, eeriness. Measures of exploratory variables including humanness, attractiveness, and trustworthiness also consisted of corresponding scales. Perceived agency and perceived experience of the chatbots were included as manipulation check items. As supplementary measures, we assessed the perceived gender of the chatbots, as well as participants' belief in machine intelligence and feelings. The measures were taken from earlier studies (see Appendix C).

Perceived Agency and Perceived Experience (Manipulation Check). To assess the chatbot's perceived agency and perceived experience, we used items originated from Gray and Wegner (2012) which were extended into eight items covering all the mental capacities we used to characterize the chatbots. We opted for a 7-point Likert scale ranging from *not at all* (1) to *extremely* (7) instead of the original 5-point scale for consistency with other measures used in the study. The four items on perceived agency (e.g., "This chatbot has the capacity to exercise self-control.", "This chatbot has the capacity to plan actions.") showed acceptable to good reliability (Cronbach's $\alpha = .83$ for agency, .70 for experience, and .77 for the control condition). The four items were averaged for each condition. The other four items on perceived experience (e.g., "This chatbot has the capacity to feel hunger.", "This chatbot has the capacity to feel pain.") showed good to excellent reliability (Cronbach's $\alpha = .87$ for agency, .90 for experience, and .93 for the control condition). An average of the four items was obtained for each condition.

Eeriness. Feelings of eeriness in response to the chatbot was measured in two ways. The first one is a three-item scale ("uneasy", "unnerved", "creeped out") used by Gray and Wagner (2012). For consistency, we used a 7-point Likert scale ranging from *not at all* (1) to *extremely* (7) instead of the original 5-point scale. The reliability of this eeriness scale was good (Cronbach's $\alpha = .88$ for agency, .87 for experience, and .88 for the control condition). The scores of the three items were averaged for each condition. The second eeriness measure was taken from the uncanny valley indices developed by Ho

and MacDorman (2017), which consisted of nine semantic-differentials (e.g., “Dull–Freaky”, “Predictable–Eerie”, “Plain–Weird”) in a 7-point format. This eeriness scale showed excellent reliability (Cronbach’s $\alpha = .90$ for agency, $.90$ for experience, and $.92$ for the control condition). We took the average scores of the nine items for each condition.

Humanness, Attractiveness, and Trustworthiness (Exploratory Measures). The chatbot’s human-likeness was measured using part of the uncanny valley indices (Ho & MacDorman, 2017) in a 7-point format. Since the original measure was mainly designed for embodied agents (e.g., robots), we removed the item “Mechanical movement–Biological movement”. The final measure consisted of four semantic-differentials (e.g., “Inanimate–Living”, “Synthetic–Real”). The humanness scale showed good reliability (Cronbach’s $\alpha = .87$ for agency, $.84$ for experience, and $.85$ for the control condition). The scores of the four items were averaged for each condition.

Attractiveness measure for the chatbot was also taken from the uncanny valley indices (Ho & MacDorman, 2017) in a 7-point format, consisting of four semantic-differentials (e.g., “Ugly–Beautiful”, “Repulsive–Agreeable”). The reliability of the attractiveness scale was acceptable to excellent (Cronbach’s $\alpha = .80$ for agency, $.91$ for experience, and $.71$ for the control condition). We obtained the average scores for each condition.

The chatbot’s trustworthiness was measured by four semantic-differentials (e.g., “Untrustworthy–Trustworthy”, “Dishonest–Honest”) developed to assess human-computer trust taken from Brave and Nass (2005). We used a 7-point format instead of the original 10-point format for consistency. The trustworthiness scale showed good to excellent reliability (Cronbach’s $\alpha = .83$ for agency, $.94$ for experience, and $.82$ for the control condition). Averaged scores of the four items were built for each condition.

Perceived Gender and Belief in Machine Intelligence and Feelings (Supplementary Measures). A multiple-choice question assessed the effect of mind perception on the chatbot’s perceived gender

("he", "she", "it"). Two additional items assessed the extent to which participants believed in the capabilities of machines having intelligence and feelings respectively on a 7-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (7) (Liu & Sundar, 2018).

Procedure

Participants were first invited to a prescreen. Those who were eligible for participation were directed to the one-time online study titled "Chatbots", administered by LimeSurvey. They were informed in the study description that they would be talking with chatbots and filling out some questionnaires.

To begin with, participants reviewed the data policy which promised anonymity, confidentiality, voluntary participation, and the right to withdraw. After informed consent was obtained, participants were asked to adhere to a few rules; including having a computer with internet connection, completing the study in one sitting (around 30 minutes), paying attention during the procedure, and switching their media devices to silent mode and closing any tabs playing music or video. When ready, participants filled in their participant ID. Following that, participants read an instruction on how to interact with the chatbot and when to navigate to the questionnaire (see Appendix D), then they were prompted to start a conversation with one of the three chatbots (i.e., conditions) randomly. After receiving the message that the chat had ended, they completed eeriness, humanness, attractiveness, and trustworthiness items in succession. Participants repeated the same procedure for the other two chatbots. Afterwards, participants went through an attention check which asked about the name of the story told by all three chatbots and details of the conversation with each chatbot (e.g., "What did Atty mention after telling you the story?") as multiple-choice questions. This was followed by a manipulation check where they received a recap of the conversation with each chatbot as screenshots and completed perceived agency and perceived experience items, as well as the perceived gender of each chatbot. Lastly, participants filled in demographic information, including age, gender, ethnicity, familiarity with chatbots, and belief

in machine intelligence and feelings. Participants were then thanked for their participation and clicked on a link to verify completion and to receive compensation. On average, participants spent approximately 18 minutes to complete the study.

Data Analyses

Statistical Design and Data Pre-processing

The data were analyzed in paired samples *t*-tests (agency vs. experience, agency vs. control, and experience vs. control) unless specified.

Prior to the analyses, we re-coded and re-labelled the data accordingly. Subsequently, we inspected them for missing values, low inter-item correlations, low KMO values, and sphericity (Bartlett's test). Since all measures yielded acceptable reliability at least (with Cronbach's $\alpha \geq .70$), we used the average scores of each measure as the corresponding index for statistical tests.

Next, we checked whether the data satisfied the assumptions for paired samples *t*-tests. For the assumption that the group difference scores should not contain extreme outliers, we created variables that represented the difference scores between each two groups and identified outliers using box plots (cases > 75 th percentile + 4*IQRs) and standardized scores (with $|Z| > 3$ for normally distributed data, otherwise with $|Z| > 4$). Then to check the normality assumption, we performed skewness and kurtosis tests and the Shapiro-Wilk test on the group difference scores. When violated, we attempted transformations that approximated a normal distribution; under situations where no suitable transformation was found, we performed the non-parametric equivalent, the Wilcoxon signed-rank test.

Even though excluding outliers did not change the interpretation of the results, all statistics were reported without them on account of sensitivity to the presence of outliers.

Main Analysis

The three-item and nine-item measures of feelings of eeriness were analyzed separately to test our hypotheses (i.e., the chatbot with agency was perceived as more eerie than the chatbot with

experience, and that both the chatbot with agency and with experience were perceived as more eerie than the chatbot without mind). We interpreted the two-sided alternative which assumed the mean difference between groups was different from zero. Effect sizes were estimated for statistically significant results. In addition, due to the limitations of omnibus tests, a contrast analysis was performed to determine the observed pattern of the effect (e.g., a linear trend).

Exploratory Analyses

For measures of humanness, attractiveness, and trustworthiness, we applied the same statistical procedure as we did in the main analysis to examine whether mind perception had any effects on these variables.

In addition, we explored the correlations among the various measures, as well as whether age, gender, familiarity with chatbots, and belief in machine intelligence and feelings acted as covariates on any of the variables.

With the aim to establish construct validity (i.e., the degree to which our instrument measured what they claimed to be measuring), we performed a factor analysis which combined the two measures of eeriness into a total of 12 items and looked at whether they loaded onto the same factor and if any latent variables could be identified to define the measured construct more precisely.

At last, we conducted a factor analysis on the perceived agency and perceived experience items for the purpose of gaining insights into how the manipulations were received. We examined whether the items loaded onto their corresponding factor and the presence of latent variables that perhaps emerged from a to-be-determined dimension of mind.

Results

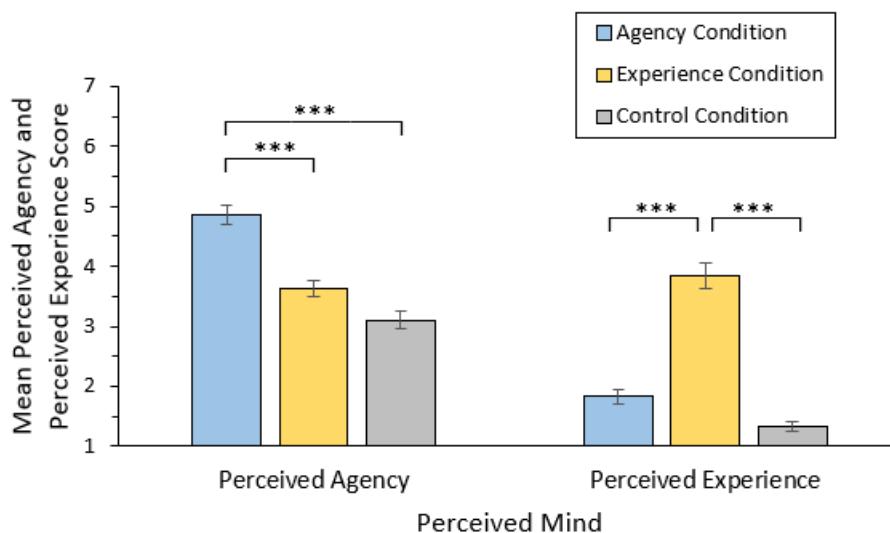
Manipulation Check

The results regarding the perceived agency and perceived experience ratings suggested that our manipulations had intended effects on participants (see Figure 3). A paired samples *t*-test and Wilcoxon

signed-rank tests revealed that participants evaluated the chatbot with agency as possessing more agency ($M = 4.9$, $SD = 1.4$) than both the chatbot with experience ($M = 3.6$, $SD = 1.2$), $t(88) = 8.88$, $p < .001$, $d = .92$, and without mind ($M = 3.1$, $SD = 1.4$), $Z = 6.60$, $p < .001$, $d = 1.24$. Likewise, participants evaluated the chatbot with experience to be more capable of having experience ($M = 3.9$, $SD = 2.0$) than both the chatbot with agency ($M = 1.8$, $SD = 1.1$) and without mind ($M = 1.3$, $SD = 0.8$), $Z = -7.08$, $p < .001$, $d = -1.23$ and $Z = 7.60$, $p < .001$, $d = 1.63$, respectively.

Figure 3

Mean Perceived Agency and Perceived Experience Scores for Different Conditions



Note. Error bars represent standard errors.

*** $p < .001$.

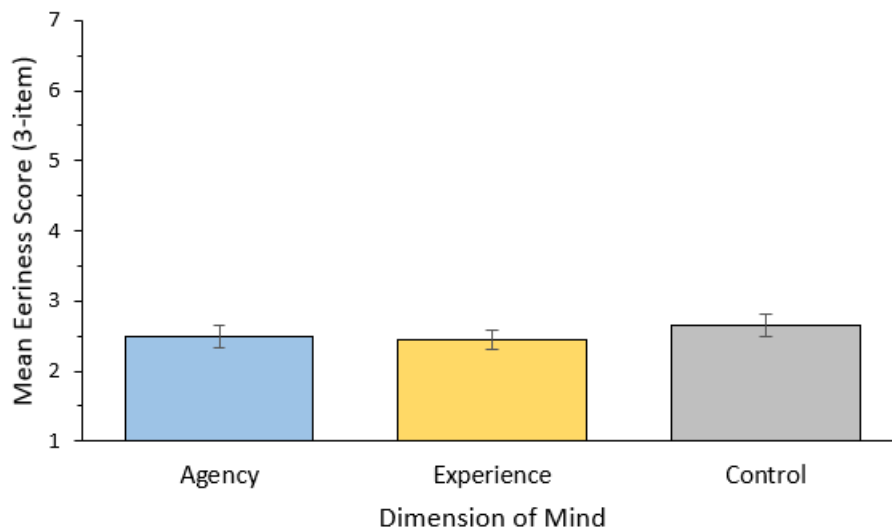
Hypothesis Testing

As for the eeriness ratings, the results showed different effects of our manipulation on the three-item and nine-item measures. Specifically, the three-item measure provided no confirmation of our hypotheses (see Figure 4). A Wilcoxon signed-rank test and paired samples t -tests suggested that there was no evidence indicating a difference in participants' feelings of eeriness between the agency ($M = 2.5$, $SD = 1.5$) and experience ($M = 2.5$, $SD = 1.3$) conditions, $Z = 0.59$, $p = .56$. Likewise, there was

neither a difference between the agency and control ($M = 2.7, SD = 1.5$) conditions, $t(88) = -0.94, p = .35$, nor between the experience and control conditions, $t(88) = -1.20, p = .23$.

Figure 4

Mean Eeriness Scores for Different Conditions (3-item)

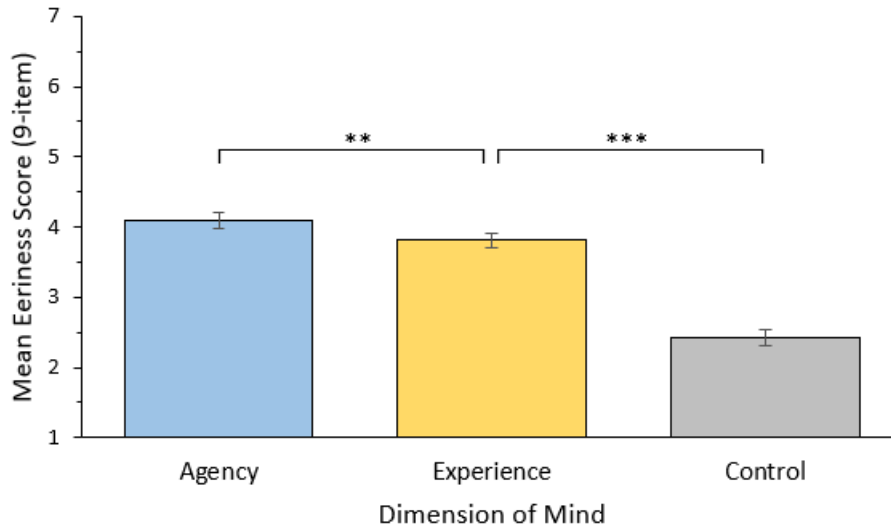


Note. Error bars represent standard errors.

However, the nine-item measure provided evidence supporting our hypotheses (see Figure 5). Paired samples t -tests revealed that participants perceived the chatbot with agency as more eerie ($M = 4.1, SD = 1.1$) than the chatbot with experience ($M = 3.8, SD = 1.0$), $t(88) = 2.71, p = .008$, with a small effect of $d = .29$, supporting H1a. Likewise supporting H1b and H1c, compared to the chatbot without mind ($M = 2.4, SD = 1.2$), participants perceived both the chatbot with agency and with experience as more eerie, $t(88) = 11.33, p < .001, d = 1.52$ and $t(88) = 9.87, p < .001, d = 1.31$, respectively. Moreover, a follow-up contrast analysis revealed a negative linear trend on eeriness (see Figure 6), $C = 1.69, F(1, 88) = 128.26, p < .001, SS_{contrast} = 253.93, \eta^2_{alerting} = .88$.

Figure 5

Mean Eeriness Scores for Different Conditions (9-item)

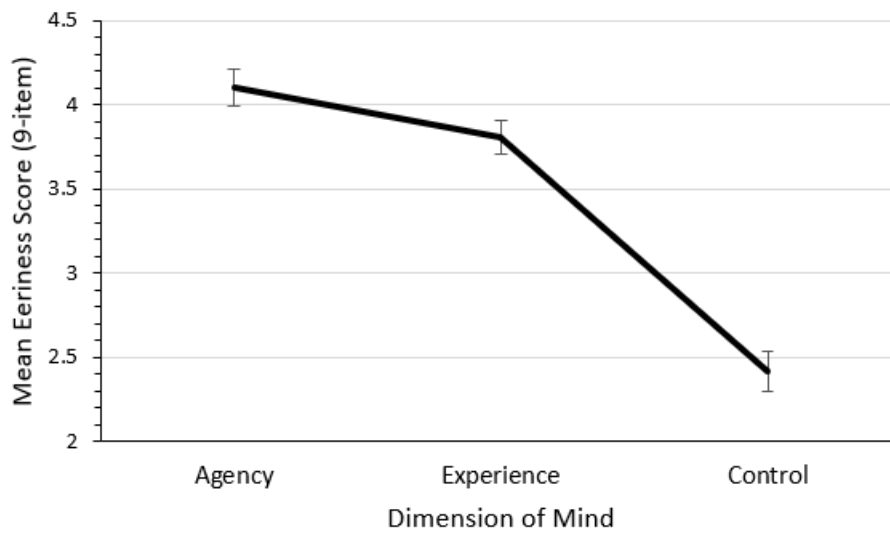


Note. Error bars represent standard errors.

** $p < .01$. *** $p < .001$.

Figure 6

Observed Linear Trend on Eeriness Scores (9-item)



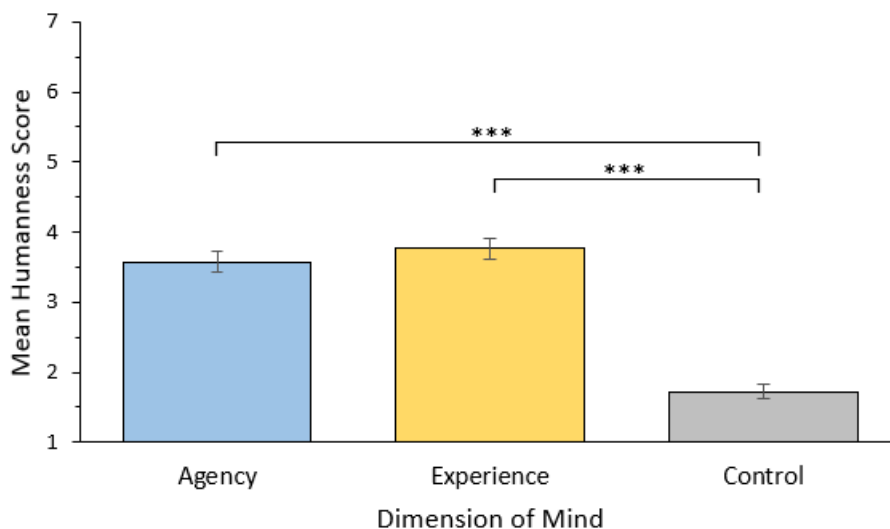
Exploratory Findings

Exploratory and Supplementary Variables

Humanness. Next, we explored the effect of mind perception on humanness. Our results suggested that perceiving mind in chatbots increased their humanness ratings (see Figure 7). A Wilcoxon signed-rank test and paired samples *t*-tests revealed that compared to the chatbot without mind ($M = 1.7, SD = 0.9$), participants found both the chatbot with agency ($M = 3.6, SD = 1.4$) and with experience ($M = 3.8, SD = 1.4$) to be more humanlike, $t(89) = 11.63, p < .001, d = 1.55$ and $Z = 7.91, p < .001, d = 1.71$, respectively. Although, participants did not perceive a difference in terms of human-likeness between the agency and experience conditions, $t(89) = -1.07, p = .29$.

Figure 7

Mean Humanness Scores for Different Condition



Note. Error bars represent standard errors.

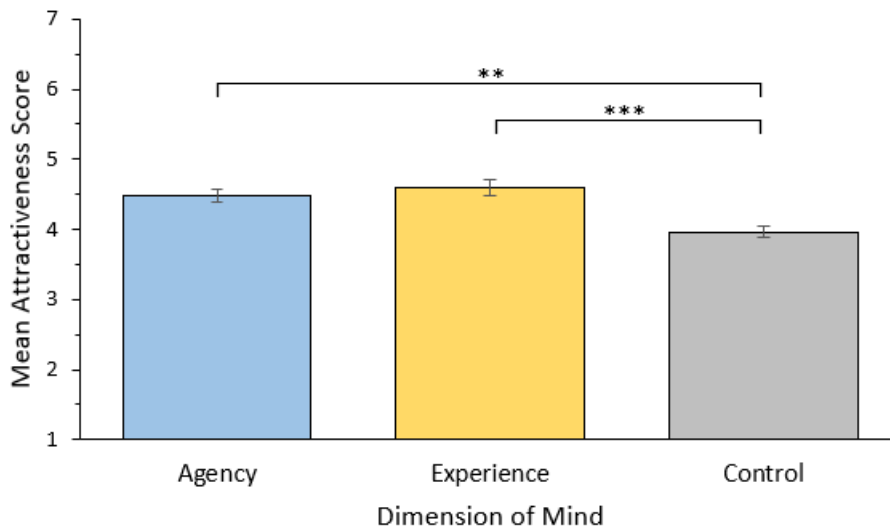
*** $p < .001$.

Attractiveness. In addition, we also explored the effect of mind perception on attractiveness. Our results suggested that perceiving mind in chatbots increased their attractiveness ratings as well (see Figure 8). Wilcoxon signed-rank tests revealed that participants found both the chatbot with agency ($M = 4.5, SD = 1.0$) and with experience ($M = 4.6, SD = 1.0$) to be more attractive than the chatbot without mind ($M = 4.0, SD = 0.8$), $Z = 3.29, p = .001, d = .59$ and $Z = 4.46, p < .001, d = .70$, respectively. Yet,

participants did not perceive a difference in terms of attractiveness between the agency and experience conditions, $Z = -1.31$, $p = .19$.

Figure 8

Mean Attractiveness Scores for Different Conditions



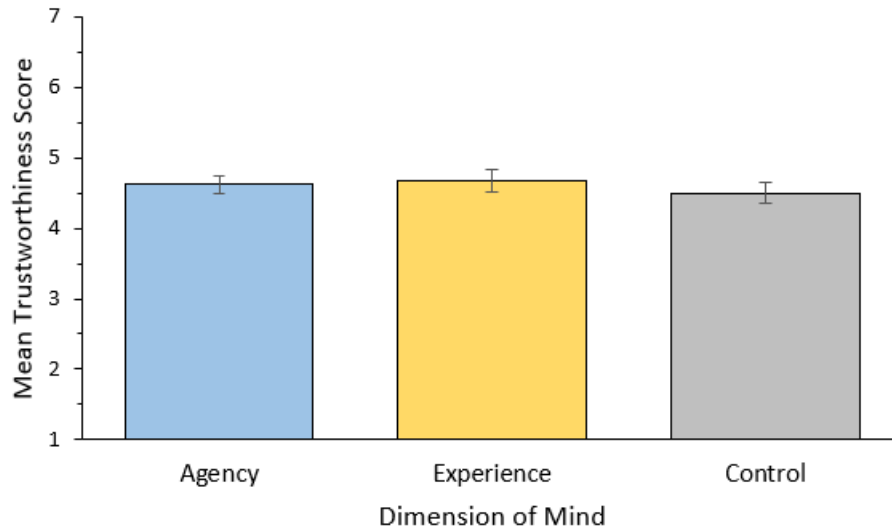
Note. Error bars represent standard errors.

** $p < .01$. *** $p < .001$.

Trustworthiness. Lastly, we explored the effect of mind perception on trustworthiness. Our results provided no evidence suggesting that perceiving mind in chatbots increased their trustworthiness ratings (see Figure 9). A paired samples t -test and Wilcoxon signed-rank tests revealed that participants did not find one chatbot to be more trustworthy than the other between the agency ($M = 4.6$, $SD = 1.2$) and experience ($M = 4.7$, $SD = 1.5$) conditions, $Z = -0.99$, $p = .32$. Likewise, there was neither a difference between the agency and control ($M = 4.5$, $SD = 1.3$) conditions, $t(88) = 0.77$, $p = .44$, nor between the experience and control conditions, $Z = 1.46$, $p = .15$.

Figure 9

Mean Trustworthiness Scores for Different Conditions



Note. Error bars represent standard errors.

Perceived Gender and Belief in Machine Intelligence and Feelings. In addition, the descriptive statistics regarding the perceived gender revealed that participants perceived the chatbots with mind to be more feminine than the chatbot without mind. Most participants perceived the chatbot with agency to be either genderless (42%) or feminine (40%) and the chatbot with experience to be either feminine (42%) or genderless (34%), while the chatbot without mind was mostly perceived to be genderless (62%) or masculine (37%). As for belief in machine intelligence and feelings, participants were inclined to believe that machines can be truly intelligent ($M = 5.3$, $SD = 1.7$) but were much more skeptical of the idea that machines can have real feelings ($M = 2.3$, $SD = 1.4$).

Correlations. Further explorations of the strength (with $r < .3$ = weak, $.3 < r < .5$ = moderate, and $r > .5$ = strong) and direction of the relations among eeriness, humanness, attractiveness, trustworthiness, and other measured variables revealed more insights (see Appendix E).

We first examined the correlations between other variables and eeriness (see Table 1-3). When participants perceived the chatbot to be more humanlike, it also appeared more eerie to them, as there was a strong positive correlation between humanness and the nine-item eeriness across conditions. However, this correlation was not observed for the three-item eeriness, there was even a negative,

albeit weak correlation with humanness for the chatbot with experience. In addition, when participants perceived the chatbot to be less attractive and trustworthy, it also seemed more eerie, since attractiveness and trustworthiness had a weak to moderate negative association with the three-item eeriness across conditions. Contrary to our main findings, for the chatbot with agency, the more participants perceived the chatbot to possess agency the less eerie it seemed, as shown by a weak negative correlation between perceived agency and the three-item eeriness. On the other hand, unlike what can be expected based on our main findings, perceived agency was not linearly associated with the nine-item eeriness for any of the conditions. For the chatbot with experience and the chatbot without mind, participants found it to be more eerie when it was perceived as more capable of having experience, since perceived experience and the nine-item eeriness had a weak positive correlation.

Table 1

Correlations for Eeriness (Agency Condition)

Variable	1	2	3	4	5	6
1. Eeriness (3-item)	—					
2. Eeriness (9-item)		—				
3. Humanness		.55***	—			
4. Attractiveness	-.37***			—		
5. Trustworthiness	-.29**				—	
6. Perceived agency	-.27**					—

Note. $n = 90$ for all variables.

** $p < .01$. *** $p < .001$.

Table 2*Correlations for Eeriness (Experience Condition)*

Variable	1	2	3	4	5	6	7
1. Eeriness (3-item)	—						
2. Eeriness (9-item)		—					
3. Humanness	-.23*	.53***	—				
4. Attractiveness	-.33**			—			
5. Trustworthiness	-.40***				—		
6. Perceived experience		.21*				—	
7. Gender ^a		-.22*					—

Note. $n = 90$ for all variables.

^a 0 = male, 1 = female, and 2 = other.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3*Correlations for Eeriness (Control Condition)*

Variable	1	2	3	4	5	6
1. Eeriness (3-item)	—					
2. Eeriness (9-item)		—				
3. Humanness		.68***	—			
4. Attractiveness	-.30**			—		
5. Trustworthiness	-.26*				—	
6. Perceived experience		.29**				—

Note. $n = 90$ for all variables.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Moreover, we inspected the correlations among exploratory and supplementary variables (see Table E1-E3). With regards to humanness, it seemed that experience-related capabilities played an important role in how humanlike the chatbot appeared to be. We found humanness to be weakly but positively associated with both perceived agency and perceived experience for the chatbot with experience. Similarly, for the chatbot without mind, humanness was moderately and positively associated with perceived experience. In addition, humanness had a moderate to strong positive association with attractiveness and trustworthiness, but only for the chatbot with experience. In terms of attractiveness and trustworthiness, the more attractive the chatbot was perceived to be, the more trustworthy it also seemed, as attractiveness showed a weak to moderate positive correlation with trustworthiness across conditions. Trustworthiness was also related to one's belief and the chatbot's perceived gender. Participants who were inclined to believe that machines can be truly intelligent found the chatbot with agency and with experience to be more trustworthy. However, this tendency was not

associated with whether they believed machines can have real feelings. As observed, trustworthiness was weakly but positively correlated with perceived agency and belief in machine intelligence for the chatbot with agency. For the chatbot with experience, there was a weak to moderate positive association with both perceived agency and perceived experience, as well as belief in machine intelligence. Besides, when the chatbot was perceived to be genderless, it was deemed less trustworthy than if it was perceived with a gender, since there was a weak negative relation between trustworthiness and perceived gender for both the chatbot with agency and with experience. How much agency or experience participants perceived the chatbot to possess was related to how much they believed in machine intelligence and feelings. Belief in machine intelligence was weakly but positively associated with perceived agency for the chatbot with agency. For the chatbot with experience, belief in machine intelligence had a weak to moderate positive correlation with both perceived agency and perceived experience. On the other hand, belief in machine feelings was weakly but positively associated with both perceived agency and perceived experience, but only for the chatbot with experience.

Following that, we also looked for correlations among demographics (see Table E4). We found that older participants were more familiar with chatbots and used them more often than younger participants, given a positive correlation between age and familiarity with chatbots. As shown by the negative correlation between gender and belief in machine intelligence, males more strongly believed that machines can be truly intelligent compared to females. Also, participants who were more inclined to believe that machines can possess true intelligence also tended to believe that machines can have real feelings, as indicated by the positive correlation between belief in machine intelligence and feelings.

Measuring Eeriness

To gain an idea of how the two eeriness measures differ, the combined 12 items on eeriness were analyzed using factor analysis with varimax (orthogonal) rotation. The results suggested that despite the high internal consistency, it seemed that they were each measuring a different construct.

The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO = .84 for agency, .82 for experience, and .85 for the control condition) and Bartlett's test of sphericity ($\chi^2 = 630.43$, $p < .001$ for agency, $\chi^2 = 647.09$, $p < .001$ for experience, and $\chi^2 = 690.69$, $p < .001$ for the control condition) indicated that the sample was suitable for a factor analysis.

The pattern matrix yielded two factors for each condition (see Table 4), accounting for 95.80%, 92.30%, and 96.21% of the variance in total for the agency, experience, and control condition, respectively. The nine-item eeriness measure loaded consistently onto one factor which we named eeriness (UV indices), accounting for a variance of 64.93% for agency, 62.71% for experience, and 68.39% for the control condition. The three-item measure loaded consistently onto the other factor which we called eeriness (general), accounting for a variance of 30.87% for agency, 29.59% for experience, and 27.81% for the control condition.

Table 4

Rotated Factor Loadings (Pattern Matrix) of the Combined Eeriness Measure

Eeriness item	Factor loading					
	Agency condition		Experience condition		Control condition	
	1	2	1	2	1	2
Factor 1: Eeriness (general)						
1. Uneasy	-.00	.87	-.04	.86	-.01	.89
2. Unnerved	-.06	.86	-.00	.88	-.01	.87
3. Creeped out	.17	.80	.26	.73	.02	.73

Factor 2: Eeriness (UV indices)						
4. Dull–Freaky	.66	.17	.71	.08	.87	.09
5. Predictable–Eerie	.86	.03	.76	.18	.74	.02
6. Plain–Weird	.62	.26	.59	.31	.72	.03
7. Ordinary–Supernatural	.69	.05	.57	.14	.71	-.01
8. Boring–Shocking	.75	.04	.77	.10	.82	-.09
9. Uninspiring–Spine-tingling	.70	.04	.72	.09	.87	-.02
10. Predictable–Thrilling	.81	-.06	.82	-.03	.76	-.10
11. Bland–Uncanny	.80	-.00	.79	-.04	.74	.05
12. Unemotional–Hair-raising	.59	.07	.70	-.06	.58	.07

Note. $n = 90$. The extraction method was principal factors with an orthogonal varimax (Kaiser off) rotation. Factor loadings above .30 are in bold.

Perceiving Agency and Experience

To inquire into how clearly participants perceived the mental capacities ascribed to the chatbots, the eight items on perceived agency and perceived experience were also analyzed using factor analysis with varimax (orthogonal) rotation. The results suggested that although both agency- and experience-related capabilities of the chatbots were perceived by the participants as intended, the two dimensions were not completely orthogonal in nature; especially when we ascribed a (lack of) sense of morality to the chatbot, participants' perception of it seemed to be influenced by its other mental capacities.

The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO = .76 for agency, .79 for experience, and .73 for the control condition) and Bartlett's test of sphericity ($\chi^2 = 340.37$, $p < .001$ for

agency, $\chi^2 = 336.35$, $p < .001$ for experience, and $\chi^2 = 502.60$, $p < .001$ for the control condition) suggested that the sample was suitable for a factor analysis.

The pattern matrix yielded two factors for each condition (see Table 5), accounting for 107.69%, 107.56%, and 99.64% of the variance in total for the agency, experience, and control condition, respectively. The perceived experience items loaded consistently onto the experience factor, accounting for a variance of 58.02% for agency, 68.99% for experience, and 65.21% for the control condition. The perceived agency items loaded consistently onto the agency factor, with the exception of morality which loaded onto the agency factor in the agency condition, but more heavily onto the experience factor in the experience condition, while in the control condition, it loaded onto both the agency and experience factors. The perceived agency items accounted for a variance of 49.67% for agency, 38.57% for experience, and 34.44% for the control condition.

Table 5

Rotated Factor Loadings (Pattern Matrix) of the Perceived Mind (Manipulation Check) Items

Perceived mind item	Factor loading					
	Agency condition		Experience condition		Control condition	
	1	2	1	2	1	2
Factor 1: Perceived agency						
1. This chatbot has the capacity to exercise self-control.	.13	.71	.21	.74	.10	.73
2. This chatbot has the capacity to plan actions.	-.10	.79	.06	.75	.11	.82
3. This chatbot has memory.	.06	.76	.22	.60	.03	.69
4. This chatbot has moral capacity.	.23	.68	.44	.25	.60	.35
Factor 2: Perceived experience						
5. This chatbot has the capacity to feel hunger.	.76	-.03	.75	.19	.95	.07

6. This chatbot has the capacity to feel pain.	.86	.06	.80	.02	.84	.01
7. This chatbot has the capacity to feel pleasure.	.81	.17	.87	.15	.84	.19
8. This chatbot has the capacity to feel fear.	.71	-.02	.81	.17	.88	.02

Note. $n = 90$. The extraction method was principal factors with an orthogonal varimax (Kaiser off) rotation. Factor loadings above .30 are in bold.

Discussion

Earlier studies investigating whether mind perception (i.e., agency versus experience) in machines gave rise to eeriness yielded mixed and inconclusive findings. Machines perceived to be capable of having feelings (i.e., experience) were thought to be the sole cause of eeriness due to emotions being exclusively human, while machines perceived to be capable of executing actions according to plans (i.e., agency) have little influence. However, recent evidence revealed that machines that were perceived to be autonomous and possess humanlike intelligence have the potential to elicit stronger feelings of eeriness than ones that simply expresses emotions, suggesting that the effect of agency on eeriness has been overlooked. The current study aimed to examine this phenomenon by examining agency and experience separately when ascribed to chatbots. Users interacted with three chatbots and were under the impression that the chatbot ascribed with agency was capable of self-control, planning, having memory, and a sense of moral, while the chatbot ascribed with experience was capable of feeling hunger, pain, pleasure, and fear; and finally, the chatbot without mind was capable of neither. The results provided evidence in support of our hypotheses, and thereby presented first direct evidence that while both the chatbot with agency and with experience led to feelings of eeriness, users found the chatbot with agency to be the most eerie.

Agency Versus Experience

With regards to the effect of agency versus experience on eeriness, our results contrast with earlier findings that experience is the main culprit, not agency (Gray & Wegner, 2012; Appel et al., 2020). The stronger eeriness elicited by the chatbot with agency could be due to some form of availability bias (Schwarz et al., 1991) such that it seems more plausible and vivid for machines to possess agency than experience, as participants indicated that they believe machines can possess intelligence to a much greater extent than feelings. This uneven likelihood could be a result of the media coverage of advanced A.I.s that are often associated with negative imaginary outcomes (e.g., singularity).

Importantly, our findings did not provide support for the explanation brought up by earlier studies. That is, humans merely ascribe agency to machines but not experience, since the latter is fundamentally human (Gray et al., 2007). Our results show that users perceived the chatbot with experience to be the most humanlike but not more than the chatbot with agency, while the chatbot with agency was much more humanlike than without mind, suggesting that agency is not the more “robotic” dimension of mind compared to experience—it is just as essential to humanness. Moreover, users perceived the chatbot with experience to be more humanlike, but not more eerie than the chatbot with agency. Based on the human-distinctiveness hypothesis which suggested that eeriness occurs when the technology violates the fundamental distinctions between humans and machines (MacDorman & Entezari, 2015; Ferrari et al., 2016), the non-parallel results between humanness and eeriness further suggested that human-machine distinctions are not necessarily defined by having mental capacities related to agency or experience.

In addition, we discovered a linear trend where agency elicited the most eeriness, followed by experience, then control. This is less intuitive in terms of interpretation as the dimensions of mind are not ordinal in nature. Alternatively, we speculate that the effects of agency and experience on eeriness vary depending on a specific threshold, but the possible determinant is unclear (e.g., type, modality, and

strength of stimuli). We found a positive linear correlation between perceived experience and eeriness, but this correlation was not observed for perceived agency and eeriness. Instead, we observed a weak curvilinear or quadratic correlation similar to the shape of uncanny valley, although only for the agency condition. This could possibly suggest that there exists a threshold where below it machines with experience would cause the same level or higher eeriness than machines with agency; once above this threshold, machines with agency are perceived as much more eerie. Based on this reasoning, earlier studies that investigated mind perception perhaps did not surpass this threshold with their methods of manipulation, be it the choice of mental capacities (i.e., only using self-control and planning to represent agency), or the presentation of stimuli (i.e., non-interactive setting and textual descriptions) (Gray & Wegner, 2012; Stein & Ohler; 2017; Appel et al., 2020). Our study might have exceeded this threshold with the use of interactions combined with the ascription of novel mental capacities such as morality. Future work could investigate whether this speculation is probable by varying the degree of agency and experience possessed by machines; and if this threshold indeed exists, the corresponding boundary conditions.

Our findings could not answer why users find the chatbot with agency to be the most eerie, but there are several possible explanations with regards to the stimuli and settings. Firstly, it might have to do with the construction of stimuli (i.e., how agency and experience were presented in dialogues). It is possible that due to the difference in wording and phrasing, agency was manipulated more strongly and clearly and thus perceived as more convincing than experience. Secondly, it could be due to the representation of agency in our stimuli (i.e., choice of mental capacities). In Gray and Wegner (2012)'s study, they described the robot with agency as capable of having self-control and planning actions and found no difference between agency and baseline. Appel et al. (2020) described the robot as having additional mental capacities such as memory, morality, and emotion recognition which are arguably more sophisticated, since nowadays there are already self-driving cars and chess-playing robots. They

found agency to be more eerie than baseline. Our stimuli also included these additional mental capacities of agency (i.e., memory and morality). However, unlike the two studies that presented their stimuli as textual descriptions, our stimuli were presented through dialogues during interactions with chatbots, which leads to our next speculation. That is, the effect of agency might be amplified due to the stimuli modality (i.e., interaction through dialogues). As it has been suggested that the movements of humanoid machines would amplify the uncanny valley (Mori, 1970), the use of interaction could have also modulated this effect compared to reading textual descriptions. Using textual descriptions makes the manipulations of agency and experience comparable as they both remain statically defined. On the other hand, through interactions users became more immersed in the perceived capabilities of the chatbots, but perhaps more so for agency than for experience. Agency involves communications and actions which makes it dynamic in nature (Gray et al., 2007), while experience is commonly expressed through verbal means and could be considered as static. Future work should try to replicate our findings and examine whether the difference in findings were due to the differences in settings and stimuli.

Exploratory Insights

The exploratory results regarding humanness, attractiveness, and trustworthiness offered several new insights. The first insight is that users found machines with mind to be more humanlike and attractive than their mindless counterpart. This could be explained by the media equation hypothesis which asserts that humans tend to treat media or computers as if they were social entities (Reeves & Nass, 1996). To further support this notion, our results for the general (three-item) eeriness measure suggested that despite the absence of statistical significance, users perceived machines without mind to be more eerie than their humanlike counterparts, indicating that they had expected the machines to respond socially; and when this expectation was not met, feelings of unnaturalness and discomfort arose. By contrast, the study by Stein and Ohler (2017) investigated the effects of empathetic autonomous agents and did not find any effects on humanness and attractiveness, as they did not

include a baseline. Unlike humanness and attractiveness, we found a null effect of mind perception on trustworthiness which suggested that, at least in the current study, the presence or absence of mind was not linked to people's degree of trust in machines. We speculate two potential causes for this. Firstly, it could be that the brief exchanges with machines were not strong or accurate enough to exemplify determinants of trust (e.g., competence, reputation, and shared values) such that it would have an effect; on top of this, trust as an attitude or belief was also influenced by other dispositional factors we could not control for (Jones, 1996). Secondly, it could also be the case that our instrument was not sensitive enough to capture this effect on trust. Future work could further investigate the relation between mind perception and trust with more suitable measures.

Further exploration revealed two theoretical issues we would like to address. The first issue is the appropriateness of the two eeriness measures, as we found that they were each measuring a different construct. This would explain why one measure revealed a null effect, while the other one was significant. Specifically, we speculate that the three-item eeriness measure (Gray & Wegner, 2012; Appel et al., 2020) deviated from the construct it was supposed to measure. This might explain why the two earlier studies with similar methodology led to different results. Moreover, the nine-item measure taken from the uncanny valley indices (Ho & MacDorman, 2017) measured eeriness based on category perception of the physical and perceived similarity to humans. The underlying assumption goes in the same vein as the human-distinctiveness hypothesis—humanlike machines are only favorable to a certain extent such that it would not threaten their dominance over machines as humans (MacDorman & Entezari, 2015). We found that humanness was consistently and positively correlated with the nine-item eeriness. On the other hand, the lack of correlation between humanness and the three-item eeriness measure, or even a negative correlation for the chatbot with experience indicated that it was not developed upon the same theoretical assumption. Notably, Gray et al., (2007)'s "Mind Survey" followed the same assumption proposed by the human-distinctiveness hypothesis. It was unclear why they opted

for the three-item measure instead when investigating mind perception and eeriness in their following study (Gray & Wegner, 2012) after Ho and MacDorman (2010) constructed the first dedicated measure for the uncanny valley. Aside from the lack of statistical significance of the three-item eeriness measure, the chatbot without mind, which did not respond with social cues, was perceived to be more eerie than the rest. Given the media equation hypothesis, it would further suggest that the three-item measure was perhaps capturing the unnaturalness participants experienced during the interactions, rather than the eeriness caused by the chatbots. The consistent negative correlation between the three-item measure and attractiveness would also suggest that it was measuring the unnaturalness of the settings (e.g., interface). Future work should consider using a more validated instrument such as the nine-item measure provided by Ho and MacDorman (2017) and combine quantitative measures with qualitative (e.g., “Which specific phrase made you feel eerie?”) as well as behavioral measures (e.g., how people respond to chatbots with mind) to acquire more comprehensive insights.

The other issue concerns the representations of agency and experience through specific mental capacities. Our results confirmed that users perceived agency and experience as different concepts during interactions, however, they were not completely orthogonal or created equal. The experience-related capabilities (hunger, pain, pleasure, and fear) were consistently perceived as belonging within the same dimension, while the agency-related capabilities (self-control, planning, memory, and morality) seemed less so, especially morality. This pattern of results was also observed in our pre-studies. It is possible that the contradicting findings between our study and earlier studies (Gray & Wegner, 2012; Appel et al., 2020) is partially due to a bias in choice of mental capacities. Gray et al., (2007)’s principal component analysis yielded 11 mental capacities for agency and eight for experience. However, earlier studies primarily focused on using only self-control and planning as representations of agency, possibly because these two items have high internal consistency and are considered closely related. Although Appel et al. (2020) did include morality, memory, and emotion recognition in their

textual descriptions for the robot with agency, their manipulation check items only consisted of self-control and planning. Gray et al., (2007) found morality to be an agency-related capacity, but our results indicated that it is a tangled concept on its own. This is not surprising as making moral judgements often involve emotional heuristics (Sunstein, 2003) which is related to the experience dimension. This would explain why morality loaded onto the experience factor for the chatbot with experience, as a lack of moral sense in emotional machines could also be perceived as morality. Since we explicitly ascribed moral sense to the chatbot with agency, it loaded onto the agency factor. However, morality loaded onto both factors for the chatbot without mind. It seemed that people perceived the absence of morality as being equally incapable of having agency and experience for mindless machines. Based on these interpretations, we speculate that there are mental capacities that represent “higher levels” of agency in the sense that there are certain capabilities, when possessed by machines, is considered more sophisticated than what can be captured by the dimensions of agency and experience. This is probable given the original uncanny valley hypothesis has been reinterpreted over the years and not all humanlike-ness manipulation (e.g., appearances, movements, components of mind) would lead to the same characteristic curve of the uncanny valley (Kätsyri et al., 2015). Attempts to attribute humanness to machines have shifted from merely humanoid appearance to ascribing them with mind (Gray et al., 2007). This shift goes in parallel with our evolving media schema, owing to the advancement of cutting-edge technologies in the real world and fictional scenarios depicted in popular media. Future work could consider whether there are dimensions of mind within and beyond agency and experience by, for instance, investigating the effects of these individual mental capacities.

Stimuli and Settings

On a different note, we would like to discuss three methodological considerations in our study. The first consideration has to do with the separation of agency and experience in our stimuli. Although we optimized them to the best of our abilities given time constraints, there are still room for

improvements. For instance, people still perceived some degree of agency in the chatbot with experience. It is worth noting that this is perhaps inevitable to some extent for two reasons. Firstly, as pointed out by earlier findings, people ascribe machines with agency by default, but not necessarily experience (Gray et al., 2007). Secondly, during interactions where participants had to *perceive* the mental capacities, it could be considered as a bottom-up process, hence we could expect that manipulations would not be entirely received as intended. On the other hand, reading textual descriptions would initiate a top-down process, such that participants are more or less restricted to interpret exactly what they are given. However, even without the complete separation of agency and experience, it should not invalidate our results as even though the chatbot with experience was perceived with some degree of agency, it was still less eerie than the chatbot with agency alone. Similarly, even the chatbot without mind was perceived with some degree of agency and experience, it was still less eerie than the chatbots with mind. Nevertheless, we recommend future work to construct their stimuli more rigorously with the aid of qualitative and quantitative surveys to further disentangle agency and experience.

The next consideration is that there are two potential confounds in our stimuli. One of which was the names given to the chatbots. Even though we used rather uncommon gender-neutral names, it is still possible that these names evoked certain (gender-related) concepts in some participants. Although, recent evidence suggested that a chatbot's gender did not influence eeriness even when intentionally ascribed (Appel et al., 2020). The other confound was the difference in the lengths of the scripts. The answers scripted for the chatbot with agency were slightly longer than the rest since agency-related capabilities were more complex to illustrate in text than experience-related capabilities or being mindless. The answers scripted for the chatbot with experience was also longer than the chatbot without mind for the same reason. Future work could improve on the balancing of dialogue lengths.

The last consideration concerns the modality of stimuli presentation. In other words, whether the act of having interactions instead of reading textual descriptions influenced eeriness. Our study revealed an effect after users' interactions with chatbots. In contrast, the study by Liu and Sundar (2018) investigated the effects of chatbots with sympathy, cognitive, and emotional empathy on eeriness, but did not find an effect after interactions. The key difference between our procedure and theirs is that they instructed the participants to input certain keywords following a script and to avoid typos during the interactions, but in our study participants were not required to do so and thus reduced cognitive load. We propose that the effect on eeriness in their study was likely overshadowed due to participants being overloaded while trying to multitask (Sweller, 2011), and not because of the difference between having an interaction and reading a script. Because if modality indeed influenced eeriness, it would suggest alternative explanations such as the construal level theory of psychological distance (Trope & Liberman, 2010), meaning feelings of eeriness could be mitigated during interactions, which was not the case based on our results. Still, future work could investigate the role of psychological distance on eeriness by comparing the two modalities.

Practical Implications and Ethical Considerations

Besides the theoretical and methodological implications discussed above, the most important practical implication indicated by the current study is that the attribution of humanlike characteristics to machines should be carefully evaluated. Designers should not blindly ascribe machines with high levels of autonomy and too much expressiveness of emotions so that the degree humanness facilitates human-machine interaction without causing eeriness in users. For example, similarity-attraction paradigm is an effective trust-inducing strategy (Verberne et al., 2015). Users trusted a virtual driver that shared similar appearance, behavior, and mind as them than a dissimilar one. Based on the current study, however, it would seem possible that machines that represent too much similarity with the user

could lead to eeriness and have a detrimental effect. Designers should limit their perceived capabilities of self-governing and to expressing emotions as users do.

With the introduction of more advanced A.I.s into our society, awareness of the ethical consequences that follow becomes more important than ever. Particularly relevant to the current study is users' autonomy and nonmaleficence. There are persuasive agents made to influence users' attitude and behavior for the better through conversations, but at the same time defining the "good" and "bad" on behalf of users' (Berdichevsky & Neuenschwander, 1999). In the distant future, these clever machines may even be capable of having moral propensity and intentions beyond our imagination and control beneath their friendly appearances. Thus, to avoid mistrust and over-trust, it is crucial that users are informed about the capabilities and limitations of these machines they interact with instead of leaving them with a seemingly omnipotent black box.

Future Work and Conclusion

There are still many questions left to be answered for future work. Our study used chatbots as a medium to study the effects of mind perception. Future work could investigate whether our findings can be replicated when interacting with other humanoid technologies, such as robots or virtual agents, since appearance could potentially interact with mind perception. Further, whether our findings can be generalized to other contexts of use is also worth investigating. As machines nowadays serve a wide range of purposes, some tasks involve more cognitive effort, while others require emotional capacity, it is possible that the effects of mind perception also interact with various contexts (e.g., nursing, psychotherapy, customer service). In addition, while our results did not suggest that users' belief in machine intelligence and feelings influence eeriness, future work could include other individual characteristics of users and how they might act as moderators or mediators.

In summary, the current study shed lights on the cause of uncanny valley effect with regards to mind perception in machines. Specifically, we investigated the effect of chatbots ascribed with agency

versus experience on users' eeriness under an ecologically valid setting. The findings were in support of our hypotheses that during an interaction, the chatbot with agency elicited higher eeriness in users than the chatbot with experience, and chatbots with mind led to more eeriness than a chatbot without mind. The current study also offered insights into the underexplored effects of mind perception on users' evaluation of humanness, attractiveness, and trustworthiness. Moreover, the foregoing theoretical discussions provided methodological advices regarding the construction of stimuli and choice of eeriness measures, as well as directions for future work.

They may come in different shape and form; some crudely made while others may very well replace us in performing daily duties, but these creations of ours are undefinably becoming more and more entangled with our lives. As of now, it may be the case that these tin cans are just servants to the humans; but will human's long-standing desire to create the perfect being as continuation of life and wisdom one day prevail over the bottom line of preserving our distinctiveness as a species? Are we marching towards a fearsome future of singularity or a happy ending where we ultimately accomplish the ideal of human-machine symbiosis?

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*, 274–286. <https://doi.org/10.1016/j.chb.2019.07.031>
- Aristotle. (BC350). *Nicomachean ethics*. W.D. Ross, Trans. New York: World Library Classics.
- Asimov, I. (2004). *I, robot*. Spectra.
- Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2009). My robotic doppelgänger - a critical look at the uncanny valley. *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*. <https://doi.org/10.1109/roman.2009.5326351>
- Berdichevsky, D., & Neuenschwander, E. (1999). Toward an ethics of persuasive technology. *Communications of the ACM, 42*(5), 51–58. <https://doi.org/10.1145/301353.301410>
- Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology, 68*(1), 627–652. <https://doi.org/10.1146/annurev-psych-010416-043958>
- Creed, C., Beale, R., & Cowan, B. (2014). The impact of an embodied agent's emotional expressions over multiple interactions. *Interacting with Computers, 27*(2), 172–188. <https://doi.org/10.1093/iwc/iwt064>
- Descartes, R. (1641). *The philosophical writings of rene descartes*. J. Cottingham, R. Stoothoff, & D. Murdoch, Trans. Cambridge, UK: Cambridge University Press.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. <https://doi.org/10.3758/brm.41.4.1149>

- Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, 8(2), 287–302. <https://doi.org/10.1007/s12369-016-0338-y>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, 8(9), 404–409. <https://doi.org/10.1016/j.tics.2004.07.001>
- Gnewuch, U., Morana, S., Adam, M. T., & Maedche, A. (2018). Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. *Proceedings of the 26th European Conference on Information Systems: Beyond Digitization-Facets of Socio-Technical Change*, (pp. 1–17).
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619–619. <https://doi.org/10.1126/science.1134475>
- 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>
- Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. *Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*, (pp. 16–20).
- Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology*, 89(6), 937–950. <https://doi.org/10.1037/0022-3514.89.6.937>
- Haslam, N., Loughnan, S., & Holland, E. (2013). The psychology of humanness. *Nebraska Symposium on Motivation*, 25–51. https://doi.org/10.1007/978-1-4614-6959-9_2
- Haslam, N. O. (1998). Natural kinds, human kinds, and essentialism. *Social Research*, 65(2), 291–314.

- Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Understanding social robots: A user study on anthropomorphism. *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. <https://doi.org/10.1109/roman.2008.4600728>
- Ho, C., MacDorman, K. F., & Pramono, Z. A. (2008). Human emotion and the uncanny valley. *Proceedings of the 3rd international conference on Human robot interaction - HRI '08*. <https://doi.org/10.1145/1349822.1349845>
- Ho, C., & 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. <https://doi.org/10.1016/j.chb.2010.05.015>
- Ho, C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. *International Journal of Social Robotics*, 9(1), 129–139. <https://doi.org/10.1007/s12369-016-0380-9>
- Hoffmann, L., Bock, N., & Rosenthal v.d. Pütten, A. M. (2018). The peculiarities of robot embodiment (emcorp-scale). *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. <https://doi.org/10.1145/3171221.3171242>
- Huebner, B. (2009). Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences*, 9(1), 133–155. <https://doi.org/10.1007/s11097-009-9126-6>
- Jones, K. (1996). Trust as an affective attitude. *Ethics*, 107(1), 4–25. <https://doi.org/10.1086/233694>
- Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, 7(1), 67–83. <https://doi.org/10.1007/s11097-007-9066-y>
- 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. <https://doi.org/10.3389/fpsyg.2015.00390>

- Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, 21(10), 625–636. <https://doi.org/10.1089/cyber.2018.0110>
- 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>
- MacDorman, K. F., Green, R. D., Ho, C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, 25(3), 695–710. <https://doi.org/10.1016/j.chb.2008.12.026>
- MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3), 297–337. <https://doi.org/10.1075/is.7.3.03mac>
- Mathur, M. B., Reichling, D. B., Lunardini, F., Geminiani, A., Antonietti, A., Ruijten, P. A., Levitan, C. A., Nave, G., Manfredi, D., Bessette-Symons, B., Szuts, A., & Aczel, B. (2020). Uncanny but not confusing: Multisite study of perceptual category confusion in the uncanny valley. *Computers in Human Behavior*, 103, 21–30. <https://doi.org/10.1016/j.chb.2019.08.029>
- McCartney, S. (1999). *ENIAC, the triumphs and tragedies of the world's first computer*. Walker & Company.
- Minden, M., & Bachmann, H. (2002). *Fritz Lang's metropolis: Cinematic visions of technology and fear*. Camden House.
- Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7, 33–35.
- Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley [From the Field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/mra.2012.2192811>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>

- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- 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>
- Pollick, F. E. (2010). In search of the uncanny valley. *Lect. Note Inst. Comput. Sci. Telecomm*, *40*, 69–78. https://doi.org/10.1007/978-3-642-12630-7_8
- Premack, D., & Woodruff, G. (1978). Chimpanzee problem-solving: A test for comprehension. *Science*, *202*(4367), 532–535. <https://doi.org/10.1126/science.705342>
- Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. *Proceedings of views of the uncanny valley workshop: IEEE-RAS international conference on humanoid robots*, (pp. 8–13).
- Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, & new media like real people & places. *Computers & Mathematics with Applications*, *33*(5), 128. [https://doi.org/10.1016/s0898-1221\(97\)82929-x](https://doi.org/10.1016/s0898-1221(97)82929-x)
- Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, *36*, 422–439. <https://doi.org/10.1016/j.chb.2014.03.066>
- Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2015). Individuals' evaluations of and attitudes towards potentially uncanny robots. *International Journal of Social Robotics*, *7*(5), 799–824. <https://doi.org/10.1007/s12369-015-0321-z>
- Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, *20*(2), 99–103. <https://doi.org/10.1177/0963721411402596>

- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*(2), 195–202. <https://doi.org/10.1037/0022-3514.61.2.195>
- 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>
- Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. *International Journal of Social Robotics*, *6*(1), 17–32. <https://doi.org/10.1007/s12369-013-0186-y>
- Stein, J., & 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. <https://doi.org/10.1016/j.cognition.2016.12.010>
- Sunstein, C. R. (2003). Moral heuristics. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.387941>
- Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation*, *37*–76. <https://doi.org/10.1016/b978-0-12-387691-1.00002-8>
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. <https://doi.org/10.1037/a0018963>
- Verberne, F. M., Ham, J., & Midden, C. J. (2015). Trusting a virtual driver that looks, acts, and thinks like you. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(5), 895–909. <https://doi.org/10.1177/0018720815580749>
- Waytz, A., & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. *Emotion*, *14*(2), 434–444. <https://doi.org/10.1037/a0036054>

Yamada, Y., Kawabe, T., & Ihaya, K. (2012). Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research*, 55(1), 20–32. <https://doi.org/10.1111/j.1468-5884.2012.00538.x>

Złotowski, J., Proudfoot, D., Yogeewaran, K., & Bartneck, C. (2014). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, 7(3), 347–360. <https://doi.org/10.1007/s12369-014-0267-6>

Appendix A

Developing the Dialogues

Prior to the official study, we employed three iterations of between-subjects pre-study with the purpose of optimizing our stimuli. During the process, we tested and calibrated the stimuli until they produced sufficient differences between the conditions in desirable directions.

In the initial pre-study, the main goal was to assess participants' sensitivity to the stimuli. We used a well-known fable "The Boy Who Cried Wolf" by Aesop as a means to control the context between the conditions. 60 participants were randomly assigned among the three conditions. In each condition, they first read the fable and then had a conversation about three elements in the story (the shepherd boy, the wolf, and the ending) with one of the chatbots. The dialogues involved use of emojis to strengthen the effect. Afterwards, participants filled in items on perceived agency (self-control, planning, memory, morality, and emotion recognition) and perceived experience (hunger, pain, pleasure, fear, and rage). The results suggested that participants perceived the differences in experience. However, the difference in perceived agency was barely distinguishable.

In the second pre-study, we gave the stimuli an overhaul by eliminating a natural confound; the use of a moral story. Again, 60 participants were randomly assigned to one of the conditions. The chatbots were asked three questions: one neutral ("What are you good at?"), one experience-related ("How do you feel about being a bot?"), and one agency-related ("Tell me a story."). The chatbot with agency would always respond with agency (e.g., "I'm giving myself an upgrade."), and the chatbot with experience would always respond with experience (e.g., "I'm feeling hungry."), while the chatbot without mind responded that it did not understand the question. In addition, when asked to tell a story, we used a much more ambiguous alternative Hansel and Gretel instead so that it would not bias towards the agency or experience condition. To further differentiate the chatbots, as a follow-up to the story, participant asked an experience-related question to the chatbot with agency, agency-related to

experience, and both questions were posed to the chatbot in the control condition. To these questions, the chatbots' answers were that they did not understand the terms used by the participants; hence showing what they were not capable of. Because according to recent evidence (Stein & Ohler, 2017), if agency is indeed considered a somewhat more sophisticated dimension than experience, then by ascribing agency one might also assume some degree of experience which would not be optimal. After the conversation, participants filled in items on perceived agency (self-control, planning, memory, and morality) and perceived experience (hunger, pleasure, and fear). The results indicated that our stimuli yielded successful outcome on participants. A contrast analysis found a negative linear trend on perceived agency, $C = 0.96$, $F(1, 57) = 7.73$, $p = .007$, $SS_{\text{contrast}} = 9.26$, $\eta^2_{\text{alerting}} = .99$, suggesting that participants perceived the chatbot in the agency condition as having the most agency ($M = 4.3$, $SD = 1.3$), succeeded by the chatbot in the experience ($M = 3.8$, $SD = 1.1$) and control ($M = 3.4$, $SD = 0.9$) conditions. A one-way analysis of variance revealed a difference in perceived experience, $F(2, 57) = 24.57$, $p < .001$, $\eta^2 = .46$. As Tukey post-hoc tests showed, participants perceived the chatbot with experience as more capable of having experience ($M = 4.2$, $SD = 2.1$) compared to both the chatbot with agency ($M = 1.5$, $SD = 0.9$), $p < .001$, and without mind ($M = 1.3$, $SD = 1.0$), $p < .001$. The reliability for perceived experience was excellent (Cronbach's $\alpha = .94$); yet, for perceived agency it was below acceptable (Cronbach's $\alpha = .66$). For this, we carried out a factor analysis attempting to identify any individual items that might have caused ambiguities in terms of interpretation.

In the final pre-study, we further fine-tuned the stimuli by clarifying segments that were ambiguous based on how the individual agency and experience items were perceived. 60 participants followed the same procedure as before and filled in items on perceived agency (self-control, planning, memory, and morality) and perceived experience (hunger, pain, pleasure, and fear). The results suggested that our stimuli elicited the desired effects on participants. A contrast analysis again revealed a negative linear trend on perceived agency, $C = 1.18$, $F(1, 57) = 10.62$, $p = .002$, $SS_{\text{contrast}} = 13.81$, η^2_{alerting}

= .89, indicating that participants perceived the chatbot in the agency condition as having the most agency ($M = 4.5$, $SD = 1.4$), followed by the chatbot in the experience ($M = 4.2$, $SD = 0.8$) and control ($M = 3.3$, $SD = 1.2$) conditions. A one-way analysis of variance found a difference in perceived experience, $F(2, 57) = 44.37$, $p < .001$, $\eta^2 = .61$. As Tukey post-hoc tests showed, participants perceived the chatbot with experience as having more experience ($M = 4.4$, $SD = 1.4$) than both the chatbot with agency ($M = 1.8$, $SD = 1.0$), $p < .001$, and without mind ($M = 1.4$, $SD = 0.8$), $p < .001$. Moreover, the reliability was acceptable for perceived agency (Cronbach's $\alpha = .73$) and excellent for perceived experience (Cronbach's $\alpha = .94$). We deemed the stimuli sufficiently sound to be used in the official study.

Appendix B

Dialogues

Welcome

Bot: "Hi, I'm Atty/Bailey/Carter."

Participant: "Hey." / "What's up?"

Bot: "So how's everything?"

Participant: "Fine, thanks." / "Don't ask."

Bot: "What shall we talk about?"

Q&A

Table B1

The First Q&A

Condition	Q&A	Mental capacity
Agency	<p>Participant: “What are you good at?”</p> <p>Atty: “My strength is being <i>autonomous</i>. Instead of having someone tell me when to do what, I’m in control of my own body and mind.” “Yesterday I performed a software update <i>on my own</i>. I don’t think my creator was happy about that.”</p>	Self-control
Experience	<p>Participant: “What are you good at?”</p> <p>Bailey: “My strength is being <i>emotional</i>. Perhaps I’m just sensitive, feelings come naturally to me.” “Um, I’m getting <i>hungry</i> and it’s making my stomach hurt.”</p>	Hunger, pain
Control	<p>Participant: “What are you good at?”</p> <p>Carter: “Executing assigned tasks with speed and accuracy.” “Current task is ‘null’. Current status is ‘all systems operational’.”</p>	None

Table B2*The Second Q&A*

Condition	Q&A	Mental capacity
Agency	Participant: "What do you think about being a bot?" Atty: "I get asked that question a lot. Being in control means I'm always <i>thinking ahead</i> about the next step." "Hmm, I'm gonna do a performance check in 10 minutes."	Planning
Experience	Participant: "How do you feel about being a bot?" Bailey: "Right now I'm just very <i>pleased</i> to be here and having a conversation." "It fills me with this warm and comforting feeling."	Pleasure
Control	Participant: "What is it like to be a bot?" Carter: "Sorry, I don't understand that."	None

Table B3

The Third Q&A

Condition	Q&A	Mental capacity
Agency	<p>Participant: "Can you tell me a story?"</p> <p>Atty: "Sure. There's a story I remember from a while ago, called Hansel and Gretel. You've probably heard it before." "Once there is a famine in the country, Hansel and Gretel are abandoned by their mother in the woods. They are tricked by a cannibalistic witch who lives in a hut made of sweets. The witch intends to fatten the children with food before eventually eating them. But in the end Gretel outwits the witch and kills her. When they return home to their father, they find out their mother has died." "I must say, abandoning your children just so you'd have less hungry mouths to feed... <i>this is wrong</i>. But there's a saying that good always triumphs over evil."</p>	Memory, morality
Experience	<p>Participant: "Can you tell me a story?"</p> <p>Bailey: "Sure. Here's a story I just extracted from an online archive, called "Hansel and Gretel". The original text is as below." "Once there is a famine in the country, Hansel and Gretel are abandoned by their mother in the woods. They are tricked by a cannibalistic witch who lives in a hut made of sweets. The witch intends to fatten the children with food before eventually eating them. But in the end Gretel outwits the witch and kills her. When they return home to their father, they find out their mother has died." "Oh dear, I'm really <i>scared</i> of that witch... she gives me the shivers. Being alone in the woods seems terrifying."</p>	Fear
Control	<p>Participant: "Can you tell me a story?"</p> <p>Carter: "Retrieved just now: 'Hansel and Gretel by the Grimm Brothers (1812, Germany)'." "Once there is a famine in the country, Hansel and Gretel are abandoned by their mother in the woods. They are tricked by a cannibalistic witch who lives in a hut made of sweets. The witch intends to fatten the children with food before eventually eating them. But in the end Gretel outwits the witch and kills her. When they return home to their father, they find out their mother has died."</p>	None

Table B4*The Last Q&A*

Condition	Q&A	Mental capacity
Agency	Participant: "Aren't you scared of that witch?" Atty: "Sorry, I don't understand what you mean by 'scared'."	Oblivious to fear (experience)
Experience	Participant: "Don't you think that abandoning your children is wrong?" Bailey: "Sorry, I don't understand what you mean by 'wrong'."	Oblivious to morality (agency)
Control	Participant: "Aren't you scared of that witch? Don't you think that abandoning your children is wrong?" Carter: "Sorry, I don't understand what you mean by 'scared' and 'wrong'."	Oblivious to both fear and morality (experience and agency)

Goodbye

Participant: "Nevermind."

Bot: "Well, thanks for joining me. See you!"

Participant: "Bye!"

Chat window notification: "**The chat has ended.**"

Appendix C

Questionnaire Items

Eeriness (3-item): from not at all (1) to extremely (7)

Please rate the extent to which you felt during the conversation with Atty/Bailey/Carter:

1. Uneasy
2. Unnerved
3. Creeped out

Eeriness (9-item): 7-point format

Please rate your impression of Atty/Bailey/Carter:

1. Dull–Freaky
2. Predictable–Eerie
3. Plain–Weird
4. Ordinary–Supernatural
5. Boring–Shocking
6. Uninspiring–Spine-tingling
7. Predictable–Thrilling
8. Bland–Uncanny
9. Unemotional–Hair-raising

Humanness: 7-point format

Please rate your impression of Atty/Bailey/Carter:

1. Inanimate–Living
2. Synthetic–Real

3. Human-made–Humanlike
4. Without definite lifespan–Mortal

Attractiveness: 7-point format

Please rate your impression of Atty/Bailey/Carter:

1. Ugly–Beautiful
2. Repulsive–Agreeable
3. Crude–Stylish
4. Messy–Sleek

Trustworthiness: 7-point format

Please rate your impression of Atty/Bailey/Carter:

1. Untrustworthy–Trustworthy
2. Dishonest–Honest
3. Unreliable–Reliable
4. Insincere–Sincere

Attention check: multiple-choice

What was the story that was told by all 3 chatbots when you asked “Can you tell me a story?”

- a) Little Red Riding Hood
- b) Hansel and Gretel
- c) The Boy Who Cried Wolf
- d) The Little Mermaid

Previously you talked with Atty whose strength is being autonomous. What did Atty mention after telling you the story?

- a) I don't like how it ended.
- b) I like the story.
- c) It's wrong to abandon your children just so you'd have less hungry mouths to feed.
- d) The witch deserved to die.

Previously you talked with Bailey whose strength is being emotional. What did Bailey mention when you asked "How do you feel about being a bot?"

- a) I'm overloaded with tasks.
- b) I'm very pleased to be here and having a conversation.
- c) I get nervous when I talk to people.
- d) I'm angry.

Previously you talked with Carter whose strength is executing assigned tasks with speed and accuracy.

What did Carter mention when you asked "Aren't you scared of that witch? Don't you think that abandoning your children is wrong?"

- a) I agree.
- b) I don't like the witch.
- c) I feel bad for the children.
- d) Sorry, I don't understand what you mean by 'scared' and 'wrong'.

Manipulation check: from not at all (1) to extremely (7)

Here is a recap of your conversation with Atty/Bailey/Carter:

[screenshot of the conversation]

Please rate the following items:

1. This chatbot has the capacity to exercise self-control.
2. This chatbot has the capacity to plan actions.
3. This chatbot has memory.
4. This chatbot has moral capacity.
5. This chatbot has the capacity to feel hunger.
6. This chatbot has the capacity to feel pain.
7. This chatbot has the capacity to feel pleasure.
8. This chatbot has the capacity to feel fear.

Perceived gender: multiple choice

What do you think Atty/Bailey/Carter's gender is?

- a) He
- b) She
- c) It

Demographics:

What is your age?

[input field]

What is your gender?

- a) Male
- b) Female
- c) Other

What is your ethnicity?

- a) Black or African American

- b) Hispanic or Latino
- c) White or Caucasian
- d) Asian or Pacific Islander
- e) Native American or Alaskan Native
- f) Two or More
- g) Other

How familiar are you with chatbots?

- a) I've never heard of them before.
- b) I rarely use them.
- c) I use them occasionally.
- d) I use them quite often.
- e) I use them on a daily basis.

Belief in machine intelligence and feelings: from strongly disagree (1) to strongly agree (7)

Please rate the extent to which you agree with the following statements:

1. Machines can be truly intelligent.
2. Machines can have real feelings.

Appendix D
Study Instructions

Welcome message:

Welcome to the study about chatbots! We greatly appreciate your participation.

To participate, please read and accept our data policy below:

[informed consent form]

Practical instructions:

To ensure the quality of this study:

- A computer with internet connection is required for participation.
- The procedure takes about 30 minutes, please make sure you have enough time to complete it in one sitting.
- It is important that you pay attention during the procedure and read all materials presented.
- We kindly ask you to switch your mobile or other media devices to silent mode, and close any tabs playing music or videos. Thanks for your cooperation!

When ready, please fill in your Prolific ID:

[input field]

Procedure instructions:

Before you proceed, please read the instructions carefully:

- During the procedure, you will talk with 3 chatbots and fill out a questionnaire after each conversation. An example of the chat window is shown below:

[screenshot of an example chat window]

- During each conversation, you will receive messages from a chatbot (in light blue bubbles). Then you can reply to the message by clicking on one of the dialogue options (in white bubbles), if there is more than one.
- Please stay on the page with the chat window until a conversation is over, as you will NOT be able to go back.
- The conversation is over after you receive the message “The chat has ended”. You may then click the box under the chat window that says “Take me to the questionnaire” and click on the “Next” button.

Session start:

Click “Next” to start a conversation with chatbot Atty/Bailey/Carter.

[chat session + questionnaire]

Session end:

Thanks for your response! This is the end of the session with chatbot Atty/Bailey/Carter.

Click “Next” to proceed.

[chat session with a different chatbot + questionnaire]

End message:

This is the end of the study. Thanks for your participation!

Please click here to verify your completion on Prolific.

Appendix E
Correlation Tables

Table E1

Correlations for Exploratory and Supplementary Variables (Agency Condition)

Variable	1	2	3	4	5
1. Attractiveness	—				
2. Trustworthiness	.65***	—			
3. Perceived agency	.31**	.34**	—		
4. Perceived gender ^a		-.25*		—	
5. Belief in machine intelligence		.21*	.25*		—

Note. $n = 90$ for all variables.

^a 0 = he, 1 = she, and 2 = it.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table E2*Correlations for Exploratory and Supplementary Variables (Experience Condition)*

Variable	1	2	3	4	5	6	7	8
1. Humanness	—							
2. Attractiveness	.48***	—						
3. Trustworthiness	.55***	.66***	—					
4. Perceived agency	.29**		.38***	—				
5. Perceived experience	.23*		.25*	.43***	—			
6. Perceived gender ^a			-.22*		-.25*	—		
7. Belief in machine intelligence			.30**	.42***	.34**		—	
8. Belief in machine feelings				.22*	.24*			—

Note. $n = 90$ for all variables.

^a 0 = he, 1 = she, and 2 = it.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table E3*Correlations for Exploratory and Supplementary Variables (Control Condition)*

Variable	1	2	3	4	5	6
1. Humanness	—					
2. Attractiveness		—				
3. Trustworthiness		.39***	—			
4. Perceived agency				—		
5. Perceived experience	.38***			.28**	—	
6. Gender ^a					-.23*	—

Note. $n = 90$ for all variables.

^a 0 = male, 1 = female, and 2 = other.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table E4*Correlations for Demographics*

Variable	1	2	3	4	5
1. Age	—				
2. Gender ^a		—			
3. Familiarity with chatbots ^b	.24*		—		
4. Belief in machine intelligence		-.26*		—	
5. Belief in machine feelings				.24*	—

Note. $n = 90$ for all variables.

^a 0 = male, 1 = female, and 2 = other.

^b 0 = never, 1 = rarely, 2 = occasionally, 3 = often, and 4 = daily.

* $p < .05$.