Stay Back, Clever Thing! Linking Situational Control and Human Uniqueness Concerns to the Aversion Against Autonomous Technology

Jan-Philipp Stein

Benny Liebold

Peter Ohler

Formal publication / citation:

Stein, J.-P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers In Human Behavior*, 95, 73–82. https://doi.org/10.1016/j.chb.2019.01.021

Acknowledgement. This study was funded by the German Research Foundation (DFG) under grant 1780 ("CrossWorlds: Connecting Virtual and Real Social Worlds"). We thank Alexandra Jost and Konrad Kosse for their help in the collection of data.

Abstract

As artificial intelligence advances towards unprecedented levels of competence, people's acceptance of autonomous technology has become a hot topic among psychology and HCI scholars. Previous studies suggest that threat perceptions—regarding observers' immediate physical safety (proximal) as well as their more abstract concepts of human uniqueness (distal)—impede the positive reception of self-controlled digital systems. Developing a Model of Autonomous Technology Threat, we propose both of these threat forms as common antecedents of users' general threat experience, which ultimately predicts reduced technology acceptance. In a laboratory study, 125 participants were invited to interact with a virtual reality agent, assuming it to be the embodiment of a fully autonomous personality assessment system. In a path analysis, we found correlational support for the proposed model, as both situational control and human uniqueness attitudes predicted threat experience, which in turn connected to stronger aversion against the presented system. Other potential state and trait influences are discussed.

Keywords: autonomous technology, artificial intelligence, threat, control, human uniqueness, virtual reality

Stay Back, Clever Thing! Linking Situational Control and Human Uniqueness Concerns to the Aversion Against Autonomous Technology

1. Introduction

In a time when people have conversations with their phones' operating systems and healthcare robots smile reassuringly while taking blood pressure, it is clear that computers have far surpassed their traditional role as passive machinery. Acknowledging this profound redefinition, a growing body of research has investigated the factors that determine the successful interaction between humans and autonomous technologies. Apart from abstract supercomputers (Gray & Wegner, 2012), self-controlled vehicles (Nordhoff et al., 2018), or sophisticated domestic robots (Bartneck et al., 2007; Kwak, Kim, & Choi, 2017), the respective studies also focus on human-like androids and virtual agents (e.g., Krämer et al., 2018; Złotowski et al., 2017), whose anthropomorphic design adds yet another factor defying traditional views on human uniqueness. Mori's *uncanny valley* model (1970) remains an influential framework in this regard, as it comprehensibly illustrates the relationship between a machine's human likeness and the way people might react to it. According to Mori's observations, human-like replicas evoke an increasingly positive response the more lifelike they appear, until a level of close—yet slightly imperfect—realism triggers intense discomfort (i.e., fear or disgust) among observers. Although several theories have been proposed to explain this phenomenon, their relative contributions remain the subject of on-going scientific debate (e.g., Shimada, Minato, Itakura, & Ishiguro, 2018; Wang, Lilienfeld, & Rochat, 2015). Moreover, due to groundbreaking innovations in the field of artificial intelligence (AI), many scholars have started to shift their focus from the "looks" of contemporary technology to its complex mental abilities, including emotional experience (Gray & Wegner, 2012) and emulated empathy (Liu & Sundar, 2018; Stein & Ohler, 2017). Now, with the realm of *uncanny minds* added to that of *uncanny appearances*, it has become more challenging than ever to disentangle the factors that cause people to feel wary in the presence of sophisticated technology.

1.1. Explanations for the Uncanniness of Technology

From the many explanations that have been suggested during nearly five decades of uncanny valley research, two main approaches can be distilled. First, an evolutionary psychological perspective has explored the role of specific perceptual cues (e.g., a robot's unrealistic eyes or movement patterns), which may prompt an aversion against pathogens (Ho et al., 2008), unfit reproductive partners (Green et al., 2008), or psychopathic individuals (Tinwell et al., 2013). Considering the innate nature of these avoidance mechanisms, the uncanny valley could indeed constitute an evolutionary phenomenon—i.e., a descendent of the inherently human fear of unfamiliarity. In a similar vein, some authors have argued that people's fear of faulty human-like machines might be related to the primal conflict that death is inevitable, assuming that such creations remind observers of the vulnerability of physical bodies (MacDorman & Ishiguro, 2006).

The second research direction, on the other hand, has focused more on the violation of overarching mental categories as an explanation for uncanny valley observations (MacDorman & Ishiguro, 2006; Yamada et al., 2013). Its proponents assume that an entity eluding previously acquired expectations (e.g., a computer acting emotionally) will cause unpleasant cognitive dissonance, triggering the impulse to avoid further contact. Remarkably, this argument is supported not only by mathematical models (Moore, 2012) but also by neuroimaging research, as fMRI studies confirm cognitive dissonance effects on a basic neurological level (Saygin et al., 2012; Urgen, Kutas, & Saygin, 2018). At the same time, it has become scientific consensus to interpret expectation violations as the product of both biological adaptation and socio-cultural influence—considering that religion, folklore, and media all contribute to people's understanding of man–machine distinctiveness and their making sense of technology (e.g., MacDorman & Entezari, 2015; Sundar, Waddell, & Jung, 2016; Young & Carpenter, 2018). Especially in Western cultures, which have been shaped by centuries of predominant Christianity, there is still an implicit tendency to regard humans as the unique "pride of creation," a species of unrivaled mental prowess and privilege (Fuller, 2014). In turn, these anthropocentric attitudes have been suspected of spawning negative views on other entities, including animals and plants, but also human-like technology (Haslam et al., 2009; Kaplan, 2004).

1.2. Autonomous Technology and Threat Experience

From a broader view, both described interpretations of the uncanny valley—evolutionary mechanism and culture-dependent categorization conflict—seem to provide quite different explanations for the aversion to advanced technology. However, it can be noted that both perspectives suggest some sort of *threat perception*, may it stem from innate or socialized factors, as the underlying cause for negative user responses. Accordingly, recent research has shown that the two approaches do indeed complement each other in the genesis of people's technology acceptance: Whereas evolutionary psychological factors predict immediate feelings of eeriness, culturally acquired attitudes might contribute indirectly by increasing people's sensitivity for the effect (MacDorman & Entezari, 2015). In consequence, scholars have hypothesized that the aversion against advanced machinery depends on both "sociocultural constructions and biological adaptations for threat avoidance" (MacDorman & Entezari, 2015, p. 141).

A similar notion can also be found in the theoretical work by Kang (2009), who anticipated control and threat perceptions as the most crucial influence on the acceptance of future technologies. Only a decade later, the autonomy of contemporary machinery has advanced far enough to turn Kang's assumptions into empirical reality; with the increasing freedom of action in modern technologies, their potential to evoke threat perceptions has grown substantially. More specifically, a recent paper by Złotowski, Yogeeswaran, and Bartneck (2017) has linked attitudes towards autonomous robots not only to perceptions of *realistic threats* (e.g., the loss of jobs, resources, and safety), but also to more symbolic *identity threats* (e.g., the loss of human uniqueness). Both of these threat categories are in fact echoed by empirical findings from other studies, although a clear differentiation does not always seem feasible. Concerning a more realistic form of threat, another experiment in the field of social robotics has demonstrated that participants experienced stronger discomfort when engaging robot groups of increased group size and coherence, as they started to expect unfavorable treatment by the mechanical "out-group" (Fraune et al., 2017). Similarly, virtual agents that threatened users' autonomy while giving environmental advice instilled strong psychological reactance among users in a previous study (Roubroeks et al., 2011). In some cases, however, observations that initially appear to involve realistic threats might also encompass more abstract, identity-related issues. For example, Waytz and Norton (2014) argued that botsourcing—the process of giving human jobs to robots—may evoke particularly strong aversion if emotion-oriented jobs are redistributed to machines, as the loss of factual resources (i.e., employment) would then be amplified by a perceived loss of identity. Similar arguments can also be found in the growing body of *dehumanization* research, which examines the significance of human uniqueness for people's self-esteem (e.g., Ferrari, Paladino, & Jetten, 2016; Haslam, 2006; Turkle, 1984; Vaes et al., 2012). Offering a comprehensive introduction to this line of thought, Biocca's article "Cyborg's Dilemma" (1997) assumes that people will feel increasingly unnatural the more they are surrounded by human-like technology; in consequence, the author suggests that human identity can only be lost while it is conquered by other, non-human entities.

Taking all of the reviewed findings into account, we note that *threat from autonomous technology* actually serves as a two-fold term in academic literature, combining concerns about immediate physical harm with the more abstract fear of delayed negative outcomes for human society. At the same time, we find that the previously established terms of realistic threats and identity threats (Złotowski, Yogeeswaran, & Bartneck, 2017) may sometimes blend with each other; as such, we suggest to reshape the dichotomy into a more flexible continuum of *threat proximity*, which ranges from threats that are very close to the physical body ("proximal threats") to those that are more immaterial and intellectual in nature ("distal threats"). Figure 1 illustrates how this novel terminology may be used to plot previous theories on a common dimension.

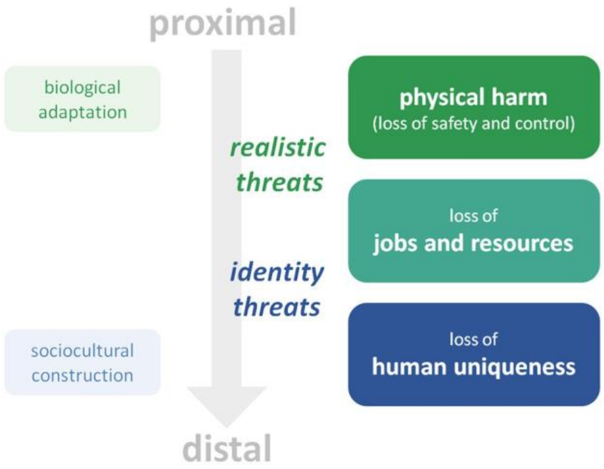

Figure 1. Threat proximity as a common dimension of previous conceptualizations (e.g., MacDorman & Entezari, 2015; Złotowski et al., 2017).

1.3. The Model of Autonomous Technology Threat

Following our review of potential threat perceptions in the face of self-controlled technology, we propose an integrative *Model of Autonomous Technology Threat* that includes (**a**) lost situational control as the prototypical proximal threat and (**b**) defied human uniqueness as the most extreme form of distal threat. Since different types of threat perception—both imminent and abstract in nature—have been shown to feed into a common neurological representation of danger (Fessler, Holbrook, & Snyder, 2012), our model proceeds to the assumption that both specific factors contribute to a more general experience of threat within the individual (Figure 2). This underlying layer, in turn, is theorized as the core predictor of people's (reduced) affinity for an autonomous technology. At the same time, we acknowledge that human uniqueness attitudes often revolve around strictly normative criteria such as religious taboos, which made us consider that the distal side of our model could also form a direct connection to the evaluation of autonomous technology, without the need to actually feel threatened.

Lastly, in terms of outcome variables, we focus on high eeriness and low attractiveness evaluations—the typical response pattern observed in empirical uncanny valley research. Since both variables have been interpreted as *emotional* qualities in previous studies (Ho & MacDorman, 2010), our model not only acknowledges the strong empirical relationship between situational control and negative affect (Rapee, 1997), but also the previously suggested link between anthropocentrism and emotional responses to technology (Nass et al., 1995).

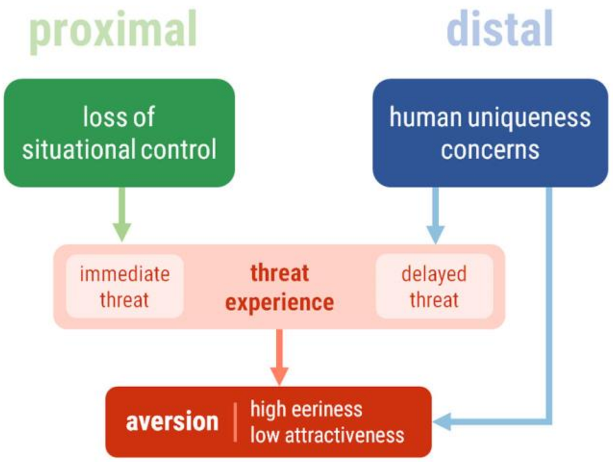

Figure 2. Model of Autonomous Technology Threat.

1.4. The Current Study

For a validation of the proposed model, we developed a laboratory study that replaced the hypothetical scenarios of previous studies with a naturalistic and interactive setting. Employing virtual reality (VR) technology, we designed a human-like virtual agent and deceptively introduced it as the embodiment of a fully autonomous personality assessment system. After a short interaction with the allegedly self-controlled technology, participants rated their situational control, concerns about human uniqueness, threat experience, and aversion. According to our model, we assumed:

H1: Less perceived situational control in interactions with autonomous technology leads to stronger aversion (proximal threat).

H2: Stronger concerns about human uniqueness lead to stronger aversion against autonomous technology (distal threat).

H3: Threat experience mediates these effects.

To increase the potential variance in the obtained data, we employed a 2×2 between-subject experimental design, striving to manipulate our model's two prototypical threat forms independently from each other. For the manipulation of situational control (as a factor contributing to *proximal threat* perceptions), we induced personal space violations, closely following the definition of interpersonal distance as a buffer zone for behavioral control (Fossataro et al., 2016; Horowitz et al., 1964; Strube & Werner, 1984). Since a pioneering study in the field of VR has indicated that digital characters can trigger similar spatial expectations as real-life encounters (Bailenson et al., 2003), we assumed that participants would experience less situational control (and thus higher proximal threat) in interactions with a physically close digital entity. To further strengthen our manipulation, we decided to present the allegedly autonomous technology in the form of a male agent, considering that personal space intrusions by male strangers have been shown to be particularly aversive to both women and men (e.g., Rustemli, 1988).

H4: Interpersonal distance violations by an autonomous technology's embodiment reduce the perceived situational control.

Addressing our model's second path, we manipulated participants' concerns about human uniqueness (as a factor contributing to *distal threat* perceptions) by presenting a specifically prepared newspaper article about the potential loss of human identity by the hands of autonomous technology. In our expectation, reading such an article before interacting with a virtual agent should (a) activate respective knowledge structures and (b) affect participants' attitudes towards autonomous technology, *biasing* them towards negative views on the provided technology. As we assumed that reading the newspaper article would also accentuate pre-existing human uniqueness concerns—which typically express themselves as attitudinal and therefore quasi-experimental factors—our manipulation ultimately strived to increase the variance within the sample.

H5: Factual biasing by a newspaper article increases concerns about human uniqueness.

Lastly—on a more exploratory note—we were interested if several other, theoretically relevant variables could offer a contribution to our model. Following our literature review, we selected the personality traits *need to belong*, which has been shown to influence basic animacy perceptions (Krämer et al., 2018; Powers et al., 2014), and *need for control*, which modulates the level of discomfort evoked by lost situational control (Leotti et al., 2010). As a contextual factor, we further looked into the *predictability* attributed to the allegedly autonomous system.

RQ: How do the need to belong, need for control, and perceived technology predictability affect participants' evaluation of an autonomous technology?

2. Method

While developing the current study, we deemed it most crucial to provide participants with the credible impression of a truly autonomous technology. In order to achieve this goal, we decided to use the Wizard of Oz method (Martin & Hanington, 2012), in which a human experimenter controls the actions of a supposedly independent agent, thereby ensuring complex but also smooth interactions. Following a 2×2 factorial design (Figure 3), participants either read a news article discussing human uniqueness concerns or not (distal threat) before interacting with a virtual agent at close or medium interpersonal distance (proximal threat).

2.1. Participants

Distributing invitation mails via Facebook groups and mailing lists, we recruited 126 undergraduate and graduate students at the local German university. Participants came from various study programs, with most of them being enrolled in media studies, engineering, and psychology. Every participant received €5 or partial course credit as a compensation for their time and effort. One participant had to be excluded from the data analysis due to not being able to correctly recall any part of the provided news article after the study. Thus, the final sample consisted of 125 participants (85 female, 39 male, 1 unspecified; *M* = 23.3 years, *SD* = 3.71). Based on our 2×2 factorial design, participants were assigned to one of four experimental conditions by means of a block randomization.

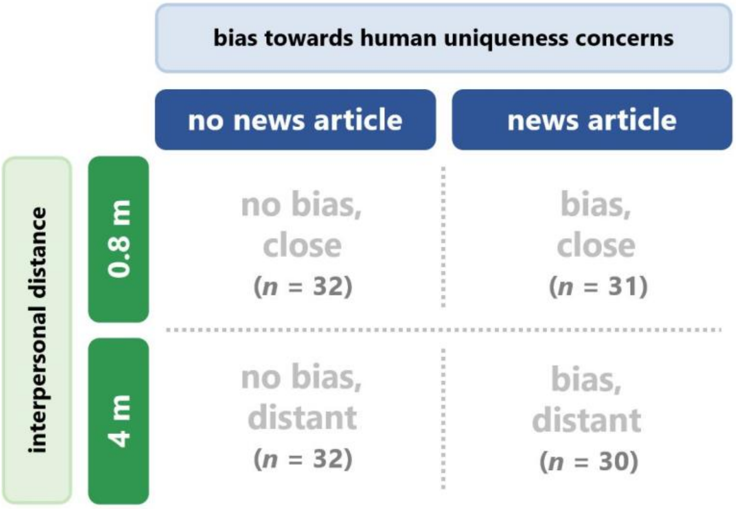

Figure 3. The study’s 2×2 between-subject design.

2.2. Procedure

Participants were invited to engage in a brief interaction with a VR agent, which we deceptively introduced as the “embodiment of a novel AI-based personality assessment system.”

Our cover story further claimed that the agent could assess anybody’s personality merely by talking to them, utilizing “a combination of voice recognition, movement tracking, word databases, and neural network technology.” To facilitate our deception, a Microsoft Kinect body tracker and microphone were assembled in our VR laboratory; in reality, a well-trained study conductor controlled all of the agent’s actions remotely, and neither the prominently placed Kinect, nor the microphone actually captured any data.

Following our introduction to the study’s (alleged) scenario, participants received a full briefing on the anonymity and voluntariness of their participation. If they had been assigned to the bias condition, they were subsequently presented with a digital newspaper article on AI technology and its potential to conquer human uniqueness; participants in the non-biased group proceeded directly to the prepared virtual environment. Having put on the provided HTC Vive head-mounted display in a comfortable manner, participants could take some time to get accustomed to the spatial orientation in the VR environment. If no discomfort occurred, we remotely activated the digital agent, making it appear on the virtual stage and approach the participant from a distance. Depending on the experimental condition, the agent either stopped at an interpersonal distance of 4 meters (distant) or came as close as 80 centimeters (close)—which is slightly less than the average personal space margin, even when accounting for culture and gender differences (Sorokowska et al., 2017). Figure 4 depicts the participants’ point of view in both conditions. Although HTC Vive hardware enables users to move around in the VR, we kindly asked participants to remain at their initial observer’s point, which was monitored by the experimenter.

Figure 4. Experimental manipulation of interpersonal distance, 4 meters (left) vs. 0.8 meters (right).

Once the agent had reached its final position, its human "wizard" controlled it to ask participants for a short personal introduction, including their three favorite hobbies. Once the respective answer had been given, the scenario always resulted in the same standardized personality analysis. Depending on the participants' talkativeness, the whole interaction lasted approximately four to five minutes, after which we asked them to complete a set of questionnaires on a laboratory PC. Concluding each appointment, participants were requested to leave their contact information so that we could inform them about the study's results—and its true nature—at a later time. If participants preferred not to leave their information, we debriefed them directly and kindly asked them to protect the integrity of our cover story in front of other students.

Stimulus Materials

We used the game engine *Unity* (Unity Technologies, version 5.5.1, 2017) to build a minimalistic virtual reality environment and the software *Adobe Fuse* (Adobe, 2017) to design a middle-aged, male virtual agent. Additionally, the *Salsa RandomEyes with LipSync* Unity plugin (Crazy Minnow Studios, 2017) was employed to synchronize the agent's lip movements with its spoken messages, and to provide realistic gaze tracking towards the participant's head position. As the core element of our Wizard of Oz deception, we coded a multi-path interaction script that enabled the study conductor to react to a variety of situations (while the participant was seemingly interacting with the autonomous system). Striving for at least moderate consistency across trials, however, the final script contained no more than 39 different speech samples. In order to give the impression of actual voice and word recognition, every interaction started with an acknowledgement of the participant's field of study ("I have understood your voice perfectly. I have not yet talked to many students from the area of [e.g., economics]!"), which was prepared in multiple versions to capture every existing faculty of the local university. Other available actions included various signs of rapport (e.g., "U-huh", "I understand"), clarifications of potential misunderstandings, and short motivational statements for cases of prolonged silence. Since we feared that a completely human voice would be disruptive for our cover story, all of the agent's spoken messages were prepared as text-to-speech sound clips using *Natural Reader* software (Naturalsoft Inc., 2017), which provides a warm, yet notably artificial voice. Lastly, we added subtle motion capturing animations (e.g., nodding, hand gestures) taken from the freely available *Virtual Human Toolkit* library (Hartholt et al., 2013) to most of the prepared statements.

According to our cover story of an advanced personality assessment AI, the interaction script had to end in some form of "analysis result." To avoid the confounding influence of different outcomes, we decided to compose a single results text, which only included ambiguous (e.g., "I have not yet decided if you are ambitious or lazy") and desirable outcomes (e.g., "I would consider you to be a rather conscientious person"), as well as statements matching the general situation of psychological experiments (e.g., “At the start of this conversation you seemed a bit shy”). Doing so, we strived to make use of the “Barnum effect” (Meehl, 1956), a phenomenon whereby vague and positive personality assessments are typically perceived as accurate by most individuals (e.g., MacDonald & Standing, 2002). Of course, to check if participants truly perceived the standardized result as authentic and personal, control questions were added to the final questionnaire.

For the bias towards human uniqueness concerns in one half of our sample, we created a detailed replica of a renowned national news website, in which we embedded a self-written article detailing how autonomous technologies “get less distinguishable from humans by the day, both in terms of cognitive and emotional capacities.” To lend a believable voice to our biasing stimulus, we mostly used excerpts from real news publications, including quotes from prominent human-computer interaction scientists and IT entrepreneurs. As a closing argument, the article claimed that “according to most experts, a clean separation of human and machine won’t be possible in the near future, resulting in countless ethical challenges.”

2.3. Measures

For the purpose of validating our Model of Autonomous Technology Threat, we needed robust measures for situational control and human uniqueness concerns, as well as checks for their respective manipulations. Due to a lack of existing instruments that could be applied to our specific scenario, we created our own questionnaires for these means; their original German versions and ad-hoc translations can be obtained from the Supplementary Materials.

Situational control. To assess their *situational control*, participants were provided with a six-item measure that consisted of both positively (e.g., “I felt as though I could react to all eventualities.”) and negatively worded (e.g., “I was at the mercy of the situation”) items, the latter of which were subsequently reverse-coded. Each item had to be filled in using a five-point response format (1 = *fully disagree*, 5 = *fully agree*). The resulting index of situational control showed acceptable internal consistency, Cronbach's $\alpha = .74$. An exploratory factor analysis further indicated that all items loaded high on a single factor, which explained 47% of the variance in our participants' answers.

Human uniqueness concerns. *Concerns about human uniqueness* were assessed with 13 self-created items (e.g., "The idea that machines will someday have the same abilities as real humans makes me anxious.") that had to be rated on a seven-point scale (1 = *fully disagree*, 7 = *fully agree*). Internal consistency of the resulting scale turned out excellent with a Cronbach's alpha of .92. While an exploratory factor analysis suggested that the questionnaire might encompass two sub-factors, all items showed a factor loading of at least .45 on the first factor, which explained 51% of the observed variance; hence, we deemed the measure valid for the assessment of human uniqueness concerns as a singular construct.

Threat experience. Participants were asked to rate their general *threat experience* as evoked by the autonomous technology with a set of ten items (e.g. "The agent was up to no good", "I was afraid of the agent") on five-point scales. Positive items such as "The agent gave a peaceful impression" were reverse-coded for our analysis. The final measure showed good internal consistency, Cronbach's $\alpha = .85$. An exploratory factor analysis indicated that the measured construct included three sub-dimensions (i.e., *kind impression* with 3 items, *fear/anxiety* with 4 items, and *suspected benevolence* with 3 items). However, a subsequent confirmatory factor analysis showed that a single second-order factor could explain between 77% and 89% of the variance in these three sub-factors, with the model's fit turning out excellent ($\chi^2 = 530.01$; $df = 45$;

$p < .001$; CFI = 0.982, TLI = 0.975, RMSEA = 0.046). Based on this, we suggest that our self-developed measure offered a sound assessment of threat experience as one overarching construct.

Technology aversion. Participants' aversion against the presented autonomous technology was assessed using the well-established uncanny valley indices by Ho and MacDorman (2010). According to the authors, the scales for perceived *eeriness* (8 items, e.g., "bland – uncanny"; Cronbach's $\alpha = .77$) and *attractiveness* (5 items, e.g., "ugly – beautiful"; Cronbach's $\alpha = .81$) constitute two distinct affective measures within the complex conceptualization of uncanniness, so that we added both as outcome variables to our analysis.

The third index of Ho and MacDorman's questionnaire, *human likeness* (6 items, e.g., "mechanical movement – biological movement"; Cronbach's $\alpha = .84$), focuses more on the uncanny valley's x-axis and was therefore included to control for potential disruptions in our groups' artificiality perceptions. However, the four conditions did not differ significantly in this regard, $F(3,121) = 0.39$, $p = .76$, meaning that we can rule out random effects in the human likeness ascribed to the presented technology.

Personal space violation and human uniqueness bias. Since we planned to manipulate the experienced situational control and human uniqueness concerns in our sample, additional measures were required to evaluate the success of both manipulations. For the induced *personal space violation*, we created seven items (e.g., "The virtual agent stood uncomfortably close to me.") in a five-point answer format (1 = *fully disagree*, 5 = *fully agree*). The resulting index proved to be of excellent internal consistency, Cronbach's $\alpha = .95$. By means of an exploratory factor analysis, we also found that all items addressed a single factor, accounting for 78% of the variance.

To make sure that our *factual biasing* in the form of a newspaper article had been read thoroughly, we composed a single-choice recall test with four questions about the provided text. By collecting answers from all participants—whether they had been biased or not—we were able to compare the biased group's recollection to chance level. Indeed, participants in the newspaper condition achieved an average of $M = 3.03$ ($SD = 1.00$) correct answers, scoring significantly higher than those in the unbiased condition ($M = 1.36$, $SD = 1.00$), Mann-Whitney $U = 3.38$, $p < .001$, $r = .64$. One participant in the bias group, who could not answer any of the four questions correctly, was excluded from our data analysis.

Additional state and trait variables. Reliable measures for participants' *need to belong* and *need for control*—the trait variables addressed by our additional RQ—could be obtained from extant literature. The need to belong scale (Leary et al., 2013) consists of ten items (e.g., "I want other people to accept me"), with a good internal consistency of Cronbach's $\alpha = .81$. The desirability of control scale (Burger & Cooper, 1979) covers 20 items regarding the individual's need for control (e.g., "I try to avoid situations where someone else tells me what to do"); since two items of the questionnaire refer to car driving, we deemed them unsuitable for our student sample and only used the remaining 18 items, observing an acceptable Cronbach's α of .71.

Technology predictability, the contextual aspect included in the exploratory RQ, was assessed with a self-created four-item scale (e.g., "I did not know what the agent would do next."; 1 = *fully disagree*, 5 = *fully agree*). Although we examined a less-than-ideal internal consistency for the measure (Cronbach's $\alpha = .61$), the emergence of a single-factor solution in a subsequent exploratory factor analysis convinced us that it could still serve its exploratory purpose.

Control variables. To identify potentially problematic outliers in our sample, we asked participants about their previous experiences with VR, video games, and virtual agents (all using one-item measures), as well as their level of *public speaking anxiety* using the well-established Personal Report of Communication Apprehension (PRCA-24; McCroskey, 1982) subscale for public speaking (6 items, e.g., “Certain parts of my body feel very tense and rigid while giving a speech”; Cronbach’s $\alpha = .82$). Neither the distribution of technological expertise nor that of public speaking anxiety indicated any notable outliers within the sample.

In the final, yet very crucial part of our questionnaire, we addressed the *plausibility* of the presented scenario with two questions. The first item explored the believability of the autonomous technology itself, asking participants to rate how competent they considered the AI system on a scale from 1 (*very incompetent*) to 5 (*very competent*). Not only did we obtain a high mean of $M = 3.92$ (SD = 0.84) for this measure, we also found no significant differences between conditions, $F(3,121) = 1.40$, $p = .25$. The second question asked participants about the perceived appropriateness of the provided personality assessment, ranging from 1 (*fully inaccurate*) to 5 (*the fully accurate*). As an a priori exclusion criterion, we decided to regard a minimum value of 1 in at least one of the two questions as a sign of overwhelming disbelief; however, no participant met this cutoff. Instead, all participants considered the system’s personality judgment to be sufficiently accurate, $M = 3.83$ ($SD = 0.68$), with no noteworthy differences between groups, $F(3,121) = 1.16$, $p = .33$. Taken together, these results highlight the more than adequate believability of our Wizard of Oz scenario across conditions.

Table 1. *Zero-order correlations between measured variables.*

	variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	eeriness	–									
2	attractiveness	__.19__*	–								
3	threat experience	.12	__-.22__*	–							
4	situational control	-.06	.08	__-.49__**	–						
5	human uniqueness concerns	.08	.07	__.20__*	-.09	–					
6	interpersonal distance violation	.10	__-.18__*	__.25__**	__-.26__**	.10	–				
7	newspaper test score (bias)	.04	-.01	.09	.04	.05	.04	–			
8	predictability	__-.43__**	-.03	-.16	.10	-.11	.01	-.13	–		
9	human likeness	.11	__.46__**	-.11	.08	-.06	.14	-.05	-.02	–	
10	need to belong	.31**	.11	.07	-.16	.26**	.00	-.08	__-.30__**	.08	–
11	need for control	-.08	-.06	.03	-.11	-.02	-.02	.16	.11	-.05	__-.21__*

Note: * $p < .05$, ** $p < .01$.

Table 2. *Means and standard deviations obtained for the measured variables.*

	low interpersonal distance				high interpersonal distance			
	no bias		bias		no bias		bias	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
eeriness[a]	4.23	0.81	4.30	0.78	4.09	0.84	4.07	0.82
attractiveness[a]	4.99	0.84	4.87	0.64	4.98	0.88	5.11	0.72
threat experience[b]	2.05	0.69	2.02	0.61	1.81	0.56	2.03	0.13
situational control[b]	2.67	0.69	2.89	0.74	3.13	0.68	2.98	0.79
human uniqueness concerns[a]	4.21	1.17	4.32	1.24	4.01	1.37	4.15	1.20
interpersonal distance violation[b]	3.63	1.02	3.89	0.94	1.50	0.48	1.68	0.54
newspaper test score (bias)[c]	33.6	27.4	76.6	21.3	34.4	22.7	75.0	28.6
predictability[b]	3.06	0.90	2.71	0.74	2.70	0.67	2.83	0.83
human likeness[a]	3.36	1.19	3.21	1.02	3.48	1.24	3.47	1.02
need to belong[b]	3.42	0.57	3.41	0.72	3.50	0.63	3.47	0.74
need for control[b]	3.46	0.33	3.57	0.37	3.44	0.37	3.50	0.44

Notes. [a] Scale range from 1 to 7. [b] Scale range from 1 to 5. [c] Percentage of correct answers.

3. Results

The threshold for statistical significance for all analyses was set to *p* < .05. Zero-order correlations for all measured variables can be obtained from Table 1. Table 2 offers an overview of the means and standard deviations of the study's measures, broken down by experimental group.

3.1. Path Analysis of the Model Variables

Based on our theoretical considerations, we conducted a path analysis using the *lavaan* package in R to find out if the correlational structure of our data conformed with our model. In this analysis, our manipulations were explored in their role as predictors of perceived situational control (H4) and human uniqueness concerns (H5), which then served as predictors for general threat experience (H1-3). In turn, threat experience was used to predict participants' technology aversion as indicated by feelings of eeriness and attractiveness. All path coefficients can be obtained from Figure 5.

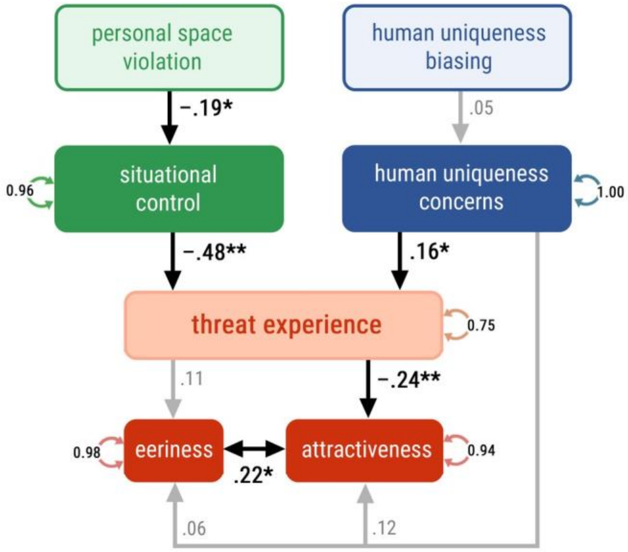

Figure 5. Coefficients obtained from path analysis (* *p* < .05, ** *p* < .01).

Although the experimental manipulation of human uniqueness concerns did not suffice to influence the underlying construct in the expected manner, the Model of Autonomous Technology Threat itself was well reflected by the obtained data. Focusing on the situational perceptions and human uniqueness attitudes by our participants, we observed both to be significant predictors of threat experience (supporting H1 and H2), which further predicted decreased attractiveness evaluations. Only the proposed direct path from human uniqueness concerns to aversion, as well as the paths to participants' eeriness perceptions turned out insignificant. Thus, we report that proximal and distal threat factors contributed to the aversion against autonomous technology as expected, albeit in a form that appears more targeted at attractiveness perceptions than at eerie feelings.

3.2. The Mediating Role of Threat Experience

A series of four mediation analyses was performed in order to investigate the potential role of general threat experience as a mediator between proximal and distal threat and the aversion against autonomous technology (H3). We used the PROCESS macro for SPSS with both eeriness and attractiveness as dependent variables and situational control and human uniqueness concerns as predictors. Indeed, the procedure revealed a significant indirect effect from situational control over threat experience on attractiveness, $b = .12$ [95% CI .01; .26], lending support for a mediation of our model's proximal component. However, no significant mediation could be uncovered for the distal path, as the impact of human uniqueness concerns on technology aversion was not mediated by a general form of threat experience.

3.3. Causal Structure

Participants indeed reported a much stronger violation of personal space when the virtual agent stood right in front of them than when it kept its distance, $t(123) = –15.48$, $p < .001$, $r = .81$.

Apart from this manipulation check, we also found that participants confronted with the close agent did in fact report significantly less situational control, $t(123) = –2.195$, $p = .03$, even though the effect turned out rather small, $r = .19$. Thus, while the manipulation of situational control can be considered successful in terms of statistical significance, we would like to give an only cautiously positive answer to H4.

On the other hand, we found no significant differences in participants’ human uniqueness concerns when comparing the group that had read the bias article compared to the group with no biasing, $t(123) = –0.569$, $p = .571$. This means that we were not able to bias participants’ attitudes towards an increased importance of human uniqueness by means of our newspaper article—indicating a negative answer to H5.

Matching the reduced effectiveness of our manipulations, a multivariate analysis of variances (MANOVA) using personal space violation scores and recall test results as independent variables and eeriness and attractiveness as dependent variables remained without significant results. Neither the manipulation of interpersonal distance violation, $V = .02$, $F(2, 120) = 1.44$, $p = .24$, nor that of human uniqueness concerns, $V < .01$, $F(2, 120) = 0.02$, $p = .98$, showed a significant isolated influence on the intensity of participants’ aversion. The interaction between both factors turned out insignificant as well, $V < .01$, $F(2, 120) = 0.49$, $p = .61$. In light of these results, we have to refrain from interpreting our findings as causal evidence for the developed model; instead, we present our analyses strictly as correlational contribution to the current literature.

3.4. Additional State and Trait Influences

To explore additional systematic variance in the measured aversion against autonomous technology, we conducted two step-wise hierarchical regression analyses including the selected exploratory variables as predictors (RQ) and either eeriness or attractiveness as the dependent variable. In both cases, all measured trait variables (human uniqueness concerns, need to belong, need for control) were entered in a first step before the measured state variables (situational control, perceived threat, agent predictability) were added in a second step.

The analysis with eeriness as criterion resulted in a significant regression equation in the first step, $F(3,121) = 4.22$, $p < .01$, explaining 7.2% of the variance in the dependent variable. Introducing state variables increased the explained variance to 18.5%, $F(6,118) = 5.69$, $p < .01$. In this extended regression model, we observed that both the need to belong, $\beta = .20$, $t(118) = 2.25$, $p = .03$, and agent predictability, $\beta = –.37$, $t(118) = -4.26$, $p < .01$, significantly predicted the eeriness ascribed to the autonomous technology: The more participants felt the need to be socially included and the less they perceived the system's actions as predictable, the eerier they rated the presented stimuli.

A second hierarchical regression with attractiveness as criterion (and the same predictor selection method) did not result in a significant regression model, neither in the first step, $F(3,121) = 0.64$, $p = .59$, nor in the second, $F(6,118) = 1.51$, $p = .18$. As such, we conclude that the additional state and trait variables had a more important influence on evaluations of the agent's eeriness, while attractiveness perceptions were more related to the factors included in our main model.

4. Discussion

The traditional understanding of the man–machine relationship has long been that of human users and obedient tools, especially in Western cultures. However, recent advances in the fields of social robotics and AI have started to contest these classic role assignments, as digital entities continue to reach unprecedented levels of autonomy. In today's technological landscape, digital minds not only make their own meaningful decisions (Banks, 2018), they also take on elaborate bodies (e.g., photorealistic VR agents) that allow for lifelike interactions. After millennia of much more restricted technology use, it comes as little surprise that people often engage these new forms of self-control and physicality with some reluctance. While studies also indicate that repeated interactions with autonomous technology might be enough to reduce users' aversion (e.g. Złotowski et al., 2016), the harmony between people and their technological creations still appears inherently fragile, raising numerous practical as well as ethical questions.

We investigated how 125 participants experienced the interaction with a realistic virtual agent, which was framed as the embodiment of an autonomous assessment system. Following the theoretical groundwork, we suspected threat experience to be the crucial mediator for potentially aversive reactions, linking both immediate (personal space invasions) and delayed threats (human uniqueness concerns) to a negative emotional response. A path analysis focusing on our hypothesized model structure lent clear empirical support to our framework. It uncovered significant paths from situational control and human uniqueness concerns to threat experience, which in turn predicted reduced attractiveness evaluations. Consequently, we note that threat perceptions—both proximal and distal in nature—indeed played an important role for the technology aversion of our participants. Moreover, a comparison of both factors' effect sizes suggests that, at least in the type of scenario that we presented, situational variables might exert a notably larger influence on threat experience than general concerns about human identity. This finding was supported by a significant mediation effect for the proximal component of our model—situational control reduced threat experience and thereby influenced the attractiveness ascribed to the virtual system. However, no significant mediation effects could be observed for distal threat factors. Thus, while human uniqueness concerns did predict threat experience (to a lesser degree), the pattern found in our data was mainly determined by situational factors. In our interpretation, this conclusion actually makes sense, as attitudinal factors shape our perceptions of a situation (e.g., to protect us from harm), but situational stimuli ultimately play a greater role in triggering actual responses. For real-life applications and the decision makers behind them, this could mean good news, considering that the situational conditions of interactive technology are mostly subject to conscious design choices. Based on our work, we argue that future technological endeavors might prove most successful in terms of user acceptance if they strive for unimpeded situational control—or, at the very least, the impression of it. Although the current study merely examined people's feelings towards a specific form of AI assessment system, our literature review and the obtained data clearly indicate that people's views on technology are modulated by stable and overarching mechanisms; thus, the shown relevance of situational factors might apply to a much larger spectrum of autonomous systems as well.

Methodologically, we adhered to the common conceptualization of uncanniness—a central outcome variable in technology acceptance research—as a blend of attractiveness and eeriness ratings. However, our path analysis showed that threat perceptions were mainly related to attractiveness, whereas eeriness relied more on other factors, such as the need to belong as a personality factor and agent predictability as a situational factor. As such, it appears that the experience of threat actually renders autonomous technology less acceptable but does not necessarily increase the strange and weird feeling reported by traditional uncanny valley experiments. In our interpretation, this finding indicates that intelligent machines falling into the "uncanny valley of mind" have reached a point beyond ambiguity and awkwardness—with their threatening gestalt merely triggering the impulse to avoid further interaction, making them appear less attractive, stylish, or beautiful.

Our participants’ perceived eeriness, on the other hand, was found to depend much more on factors such as the personality trait *need to belong*: The more people valued social inclusion, the more they disliked the presented AI system. We suspect a possible cause for this in the lack of interpersonal warmth during our prepared interaction scenario. Since we had scripted the agent to present a highly ambiguous personality judgement (including the rude statement “I have heard more interesting hobbies than yours”), its actions might have felt like a form of rejection to some participants. We assume that those with a higher need to belong may have been especially surprised or hurt by the analytic behavior of the assessment system, culminating in the impression of an eerie, not entirely normal human-computer interaction.

Finally, a particularly strong predictor ($\beta = –.37$) for eeriness emerged in the form of *technology predictability*, which we had initially conceptualized as a sub-factor of situational control. In light of this, we come under the impression that behavioral predictability may provide a much more fitting construct to explain uncanniness than situational control in general; unlike the latter, predictability focuses exclusively on the stimulus itself, setting aside internal attributions (such as self-efficacy) and other interfering situational variables. Considering that predictability has also been shown to be an important antecedent of threat perceptions on a neurobiological level (e.g., Klahn et al., 2016), the concept might be another important construct to be included in our Model of Autonomous Technology Threat. Therefore, we suggest a potential model extension for future applications of our work, although a more refined measure for technology predictability might be needed in this case.

Limitations

Familiarity, likability, affinity—uncanny valley researchers often disagree about a meaningful conceptualization of the y-axis in Mori’s model, which he originally described with the Japanese neologism “shinwakan.” Complicating matters even further, several authors have argued that the concept is actually multi-dimensional (MacDorman & Ishiguro, 2006; Moore, 2012), raising doubts on the idea to simply measure technology aversion with a single variable. In consequence, it has become common practice to assess people’s disliking of human-like technology with multiple concepts, such as the presently used, two-fold operationalization as high eeriness and low attractiveness. According to Ho and MacDorman (2010), who developed a widely used instrument to measure these two variables, this conceptualization is actually supported by high reliability coefficients and an only marginal intercorrelation. Indeed, our data showed that both dependent variables were shaped by quite different predictors, as some factors exerted a stronger influence on eeriness (need to belong, predictability) and others were more related to attractiveness evaluations (threat experience). At first sight, these results indicate that the uncanny valley’s current operationalization provides the complexity needed to describe affective responses to sophisticated technology. However, we think that the underlying mechanisms still remain oversimplified by this type of measurement, as uncanniness encompasses a multitude of affective (e.g., fear, disgust), physiological, and behavioral facets. At least in regard to autonomous technology, it could therefore be considered to examine all of these indicators separately than to summarize them under an umbrella term.

During our effort to increase the natural variance of both types of threat perception in our sample, we encountered some difficulties, especially when trying to influence human uniqueness concerns by means of a newspaper-based bias stimulus. Unfortunately, our observation conforms with a recent critique about the many contingencies of related procedures (Cesario, 2014), indicating that attitudes cannot be sufficiently manipulated by short-term interventions. While the manipulation of situational control turned out more successful, the influence of our personal space violations was also not as strong as expected. Therefore, the evidence we provide for our model is correlational in nature. Additional experimental work focusing on the importance of threat-related factors for the aversion against autonomous technology could further strengthen this line of research.

While the natural interaction paradigm used in our experiment underscores the high ecological validity of our findings, we still have to note several limitations due to the use of a convenience sample of local university students. Specifically, both the similar level of education among our participants, as well as their homogenous age arguably lessens the generalizability of our results. In regard to our theoretical framework, we further note that the relatively high percentage of atheists within our sample (69.6% claimed to be unreligious) has likely weakened the impact of human uniqueness concerns—an attitudinal construct that typically revolves around religious views and might even require a certain level of fundamentalism (MacDorman & Entezari, 2015). In light of this, we assume that our data might underestimate the distal path of the developed model, and that worries about human distinctiveness should emerge as a much stronger influence among other samples, especially those rooted in anthropocentric beliefs (e.g., conservative Christians, Muslims).

5. Conclusion

The current study set out to explore the role of perceived threat in interactions with autonomous technology. We juxtaposed two pathways that might fuel a general level of threat experience: Proximal threats that build upon the behavior of the respective technology and the surrounding environment, and distal threats that result from abstract ruminations about human identity. In a laboratory experiment, we found first support for the proposed model—but also inspiration for necessary modifications. Compared to studies that use hypothetical scenarios, our results achieved higher external validity as they emerged from an actual interaction with what was believed to be a fully autonomous technology. At the same time, the reported evidence is mostly correlational in nature; also, to our surprise, only attractiveness ratings were influenced by participants' experience of experience, whereas the eerie impression of "getting goosebumps" depended on more specific factors such as the technology's predictability. In summary, more pronounced manipulations of proximal and distal threats, as well as new ways of measuring technology aversion might be necessary to advance this fascinating research area. Eventually, design implications derived from this line of research could help to ensure that the potential benefits of autonomous technology are not lost because its users feel threatened, replaced, or hopelessly out of control.

References

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin, 29*(7), 819–833. https://doi.org/10.1177/0146167203253270

Banks, J. (2018). A perceived moral agency scale: development and validation of a metric for humans and social machines. *Computers in Human Behavior, 90,* 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Bartneck, C., Verbunt, M., Mubin, O., & Al Mahmud, A. (2007). To kill a mockingbird robot. In *Proceedings of the 2007 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 81–87). ACM Press. https://doi.org/10.1145/1228716.1228728

Biocca, F. (1997). The cyborg's dilemma: Embodiment in virtual environments. *Journal of Computer-Mediated Communication, 3*(2). https://doi.org/ 10.1111/j.1083-6101.1997.tb00070.x

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion, 3*(4), 381–393. https://doi.org/10.1007/BF00994052

Cessario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science, 9*(1), 40–48. https://doi.org/10.1177/1745691613513470

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

Fessler, D. M., Holbrook, C., & Snyder, J. K. (2012). Weapons make the man (larger): Formidability is represented as size and strength in humans. *PLOS One, 7*(4): e32751.

Fossataro, C., Sambo, C. F., Garbarini, F., & Iannetti, G. D. (2016). Interpersonal interactions and empathy modulate perception of threat and defensive responses. *Scientific Reports*, *6*, 19353. https://doi.org/10.1038/srep19353

Fraune, M. R., Nishiwaki, Y., Šabanović, S., Smith, E. R., & Okada, M. (2017). Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 205–213). ACM Press. https://doi.org/10.1145/2909824.302024

Fuller, M. (2014). The concept of the soul: Some scientific and religious perspectives. In M. Fuller (Ed.), *The Concept of the Soul: Scientific and Religious Perspectives* (pp. 1–4). Newcastle upon Tyne: Cambridge Scholars Publishing.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*, 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Green, R. D., MacDorman, K. F., Ho, C. -C., & Vasudevan, S. K. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior, 24*(5), 2456–2474. https://doi.org/10.1016/j.chb.2008.02.019

Hartholt, A., Traum, D., Marsella, S. C., Shapiro, A., Stratou, G., Leuski, A., ... Gratch, J. (2013). All together now: Introducing the Virtual Human Toolkit. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaira (Eds.), *Proceedings of the 13th International Conference on Intelligent Virtual Agents* (pp. 368–361). Berlin/Heidelberg: Springer.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003/4

Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2009). Attributing and denying humanness to others. *European Review of Social Psychology, 19*, 55–85. https://doi.org/10.1080/10463280801981645

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior, 26*, 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Ho, C.-C., MacDorman, K. F., & Pramono, Z. A. D. (2008). Human emotion and the uncanny valley: A GLM, MDS, and ISOMAP analysis of robot video ratings. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction*, 169–176.

Horowitz, M. J., Duff, D. F., & Stratton, L. O. (1964). Body-buffer zone. Exploration of personal space. *Archives of General Psychiatry*, *11*(6), 651–656. https://doi.org/10.1001/archpsyc.1964.01720300081010

Kang, M. (2009). The ambivalent power of the robot. *Antennae*, *1(9)*, 47–58.

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*, *1(3)*, 465–480.

Klahn, A. L., Klinkenberg, I. A., Notzon, S., Arolt, V., Pantev, C., Zwanzger, P., & Junghöfer, M. (2016). Prepare for scare—Impact of threat predictability on affective visual processing in spider phobia. *Behavioural Brain Research, 307*, 84–91. https://doi.org/10.1016/j.bbr.2016.03.045

Krämer, N. C., Lucas, G., Schmitt, L., & Gratch, J. (2018). Social snacking with a virtual agent—On the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *International Journal of Human-Computer Studies*, *109*, 112–121. https://doi.org/10.1016/j.ijhcs.2017.09.001

Kwak, S. S., Kim, J. S., & Choi, J. J. (2017). The effects of organism- versus object-based robot design approaches on the consumer acceptance of domestic robots. *International Journal of Social Robotics, 9*(3), 359–377. https://doi.org/10.1007/s12369-016-0388-1

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct validity of the need to belong scale: Mapping the nomological network. *Journal of Personality Assessment, 95*(6), 610–624. https://doi.org/10.1080/00223891.2013.819511

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origin and value of the need for control. *Trends in Cognitive Sciences, 14*(10), 457–463. https://doi.org/10.1016/j.tics.2010.08.001

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). https://doi.org/doi.org/10.1089/cyber.2018.0110

MacDonald, D. J., & Standing, L. G. (2002). Does self-serving bias cancel the Barnum effect? *Social Behavior and Personality, 30*(6), 625–630. https://doi.org/ doi:10.2224/sbp.2002.30.6.625

MacDorman, K. F., & Entezari, S. (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., & Ishiguro, H. (2006). The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies, 7*(3), 297–337. https://doi.org/10.1075/is.7.3.03mac

Martin. B., & Hanington, B. (2012). *Universal methods of design* (pp. 204–205). Beverly: Rockport Publishers.

McCroskey, J. C. (1982). *An introduction to rhetorical communication*. Englewood Cliffs, NJ: Prentice-Hall.

Meehl, P. (1956). Wanted—a good cookbook. *American Psychologist, 11*(6), 263–272. https://doi.org/ 10.1037/h0044164

Moore, R. K., (2012). A Bayesian explanation of the ‘uncanny valley’ effect and related psychological phenomena. *Scientific Reports, 2*(864), 1–5. https://doi.org/10.1038/srep00864

Mori, M. (1970). The uncanny valley. *Energy*, *7(4)*, 33–35.

Nass, C. I., Lombard, M., Henriksen, L., & Steuer, J. (1995). Anthropocentrism and computers. *Behaviour & Information Technology*, *14(4)*, 229–238.

Powers, K. E., Worsham, A. L., Freeman, J. B., Wheatley, T., & Heatherton, T. F. (2014). Social connection modulates perceptions of animacy. *Psychology Science*, *25*(10), 1943–1948. https://doi.org/10.1177/0956797614547706

Rapee, R. M. (1997). Perceived threat and perceived control as predictors of the degrees of fear in physical and social situations. *Journal of Anxiety Disorders, 11*(5), 445–461. https://doi.org/10.1016/S0887-6185(97)00022-4

Roubroeks, M., Ham, J., & Midden, C. (2011). When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. *International Journal of Social Robotics, 3*(2), 155–165. https://doi.org/10.1007/s12369-010-0088-1

Rustemli, A. (1988). The effects of personal space invasion on impressions and decisions. *The Journal of Psychology, 122*(2), 113–118.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7(4)*, 413–422. https://doi.org/10.1093/scan/nsr025

Shimada, M., Minato, T., Itakura, S., & Ishiguro, H. (2018). Uncanny valley of androids and the lateral inhibition hypothesis. In H. Ishiguro & F. Dalla Libera (Eds.), *Geminoid Studies* (pp. 137–155). Singapore: Springer.

Sorokowska, A., Sorokowsi, P., Hilpert, P., Cantarero, K., Frackowiak, T., Ahmadi, K., ... Pierce, J. D. (2017). Preferred interpersonal distances: A global comparison. *Journal of Cross-Cultural Psychology, 48*(4), 557–592. https://doi.org/10.1177/0022022117698039

Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, *160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Strube, M. J., & Werner, C. (1984). Personal space claims as a function of interpersonal threat: The mediating role of need for control. *Journal of Nonverbal Behavior, 8*(3), 195–209. https://doi.org/10.1007/BF00987291

Sundar, S. S., Waddell, T. F., & Jung, E. H. (2016). *The Hollywood robot syndrome: Media effects on older adults' attitudes toward robots and adoption intentions*. Paper presented at the 11th Annual ACM/IEEE International Conference on HumanRobot Interaction, Christchurch, New Zealand.

Tinwell, A., Abdel Nabi, D., & Charlton, J. (2013). Perception of psychopathy and the uncanny valley in virtual characters. *Computers in Human Behavior, 29*(4), 1617–1625. https://doi.org/10.1016/j.chb.2013.01.008

Turkle, S. (1984). *The second self. Computers and the human spirit*. New York: Simon & Schuster.

Urgen, B. A., Kutas, M., & Saygin, A. P. (2018). Uncanny valley as a window into predictive processing in the social brain. *Neuropsychologica*, *114*, 181–185. https://doi.org/10.1016/j.neuropsychologia.2018.04.027

Vaes, J., Paladino, M. P., Castelli, L., Leyens, J. P., & Giovanazzi, A. (2003). On the behavioral consequences of infra-humanization: The implicit role of uniquely human emotions in intergroup relations. *Journal of Personality and Social Psychology*, *85*(6), 1016–1034. https://doi.org/10.1037/0022-3514.85.6.1016

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*(4), 393–407. https://doi.org/10.1037/gpr0000056

Waytz, A. & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking – not feeling – jobs. *Emotion*, *14*, 434–444. https://doi.org/10.1037/a0036054

Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research*, *55*, 20–32

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction*, *5*(2), 29–47. https://doi.org/10.5898/JHRI.5.2.Yogeeswaran

Young, K. L., & Carpenter, C. (2018). Does science fiction affect political fact? Yes and no: A survey experiment on “Killer Robots”. *International Studies Quarterly*, *62*(3), 562–576. https://doi.org/10.1093/isq/sqy028

Złotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology, 6,* 883. https://doi.org/10.3389/fpsyg.2015.00883

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies, 100,* 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008