

More Human Than Human: Does The Uncanny Curve Really Matter?

Jakub Złotowski
HIT Lab NZ
University of Canterbury
Christchurch, New Zealand
jakub.zlotowski@pg.canterbury.ac.nz

Diane Proudfoot
Philosophy and Political Science
University of Canterbury
Christchurch, New Zealand
diane.proudfoot@canterbury.ac.nz

Christoph Bartneck
HIT Lab NZ
University of Canterbury
Christchurch, New Zealand
christoph.bartneck@canterbury.ac.nz

Abstract—Anthropomorphism is a common phenomenon known already in ancient times. It is not a thing of the past, but still has a profound impact on major aspects of our lives and on research in AI and HRI. Its importance in the field of HRI is emphasized by the hotly-discussed uncanny valley hypothesis. However, in spite of its popularity, the uncanny valley hypothesis lacks empirical evidence. In this paper we suggest that the community should stop trying to fit data to this hypothesis, but rather, based on the available evidence, start talking about the ‘uncanny curve’. Moreover, we point out mistakes in the previous studies of the uncanny curve and strongly encourage exploring it in a real HRI for it to be really relevant. We suggest that understanding the opposite process of anthropomorphisation, known as dehumanization, can help to cross the uncanny bottom of the graph.

Keywords—human-robot interaction; uncanny valley; uncanny curve, anthropomorphism.

I. INTRODUCTION

The word *anthropomorphism* has ancient Greek origins and consists of two parts: *anthrōpos*, meaning man or human and *morphic*, having a specific shape or form. A phenomenon of people attributing human-like characteristics to their surrounding environment was described for the first time by Xenophanes. In the 6th century BC he noted that the gods whom people worship are depicted to resemble their believers [1]. He further suggested that if horses or oxen had hands, they would draw figures of gods that look similar to horses or oxen. The importance of anthropomorphism in shaping religious beliefs proposed by Xenophanes remained central in modern theories of religion [2].

The phenomenon of anthropomorphism is not only timeless, but also widespread though various domains that affect human life, behaviour and laws. Anti-choice advocates liken an unborn human fetus to a human being as one of the main arguments opposing abortion [3]. In Spain captive chimpanzees were granted limited human rights as a result of evidence for the presence of mind [4]. Moreover, it is common to call our planet ‘Mother Earth’ and anthropomorphism is involved in discussions about environmental concerns [5]. It is used to sell products [6], [7] or design user-friendly technological agents [8]. Finally, it has been used to describe non-human animals [9], weather patterns [10] or moving geometrical figures [11].

A. Why do we anthropomorphize?

Many recent theories of religion make stronger claims about the role of anthropomorphism in religion, originating with the philosopher David Hume, who famously said, ‘There is an universal tendency among mankind to conceive all beings like themselves ... We find human faces in the moon, armies in the clouds’ [12, p. 29]. These theories hypothesize that an evolved tendency to anthropomorphism explains both the origin and persistence of religious experience and beliefs (e.g. [13]–[17]). Anthropomorphism is also involved in certain auditory and visual hallucinations associated with epileptic or psychiatric disorders – for example, the ‘sensed presence’ (or ‘felt presence’) phenomenon, where a subject senses or feels another person’s presence, when in fact no one is there (e.g. [18]). The sensed-presence phenomenon can be induced by stimulating electrodes implanted in the left temporoparietal junction of the brain [19]. Neuroscientists pinpoint different areas of the brain as the neural correlates of religious and similar anomalous experiences (e.g. [20]); the temporal lobe is frequently suggested as the location of religious experiences (e.g. [21], [22]). Researchers have also found a correlation between the sensed-presence phenomenon and certain personality characteristics, including suggestibility [23].

If we have evolved to see and hear humans and human-like gods so readily, what explains this tendency? Based on Guthrie [2], various theorists have argued that anthropomorphism is adaptive; early humans who interpreted ambiguous shapes as human minimized their risks of being killed by enemies and maximized their chances of making friends. A special-purpose, hair-triggered mechanism to detect agents – the ‘hypersensitive agency detection device’ (HADD) – has been hypothesized (e.g. [24], [25]). It is also argued that several different psychological mechanisms generate the impression of agents [26].

II. ANTHROPOMORPHISM IN AI

A. The centrality of anthropomorphism in AI

Computer scientists have long been aware of how easily humans anthropomorphize machines. In 1948 Turing said that playing chess against even a ‘paper machine’ (a simulation of machine behaviour by a human being using paper and pencil) gives ‘a definite feeling that one is pitting one’s wits against something alive’ ([27, p. 412]; see [28], [29]). Researchers have varying aims in building anthropomorphic robots to be

used in the investigation of human-robot interaction. Some social roboticists have narrow aims. These include using HRI in order to: test psychological hypotheses about human social and cognitive behaviour and development (e.g. [30]); produce service, entertainment, and therapeutic robots, with which humans with no specialized training can interact intuitively (e.g. [31]–[34]); increase learning and training opportunities for machines, by building machines that can learn new behaviours from humans via normal social cues (e.g. [35]). The ‘believability’ of the robot is particularly important in socially assistive robotics (see [36]; for the notion of ‘believable creatures’, see [37]).

Roboticists may also have grander aims. Anthropomorphism is central to AI in ways that go to the philosophical foundations of the field. Turing suggested in 1950 that one approach to machine intelligence would be to provide a machine with ‘the best sense organs that money can buy’, and then ‘teach it to understand and speak English’ [38, p. 460]. The humanoid robot has often been seen as the ‘holy grail’ of AI. More recently proponents of embodied and socially situated AI have argued that intelligence requires embodiment, and that human-like intelligence requires human-like embodiment. Descartes notoriously said that, even if machines looked like human beings and ‘imitated our actions as closely as possible for all practical purposes, we should still have certain means of recognizing that they were not real men’; they would not be able to use language ‘so as to give an appropriately meaningful answer to whatever is said in its presence, as the dumbest of men can do’ and they would behave in ways that ‘would reveal that they were acting not through understanding but only from the disposition of their organs’ [39, pp. 139-140]. Modern work in AI on anthropomorphic machines, and on human-robot interaction that resembles human-human interaction, is a potential reply to such scepticism.

B. Is there nothing to ‘mind’ but anthropomorphizing?

It has been suggested that caregivers naturally and unwittingly anthropomorphize infants – for example, by interpreting a neonate’s purely reflexive facial display as a smile. Anthropomorphizing human infants, by attributing intentions and affect to babies, may have evolutionary value: it promotes infant-caregiver bonding and enables the infant to learn social interactions. Social and developmental roboticists aim to exploit the same tendency to anthropomorphism, in order to build ‘socially intelligent’ robots (see e.g. [40], [41]). These researchers are attempting to build what Turing called a child-machine ([38]; see [42]). They emphasize ‘theory of mind’ abilities, such as face and agent detection, joint visual attention, self-recognition, and success in false belief tasks (e.g. [43]–[46]). These abilities are acquired early in development and serve as the building-blocks for later cognitive behaviour.

Researchers aim to exploit theories in human ethology and developmental psychology, creating a system architecture for the robot that is analogous to a human’s (hypothesized) cognitive architecture (e.g. [47]). Human observers interact with such robots (the canonical example is the now-retired Kismet) in some ways as if the robots were infants. The researchers’ aim is also that the robot will learn just as the infant learns, via the observer’s anthropomorphizing. This raises a critical philosophical question: if we must anthropomorphize both

humans and machines, and both as a result acquire ‘social intelligence’, why deny that machines can think?

C. The risks of anthropomorphism in AI

The focus on human-like AIs has been criticized for various reasons (see [48]). It is sometimes argued that an emphasis on anthropomorphic machines leads the general public to misunderstand the aims and achievements of AI (e.g. [49]). Powerful voices within AI have argued for research into ‘generic’ intelligence, and against the idea of imitating human performance (e.g. [50]).

The real danger to the field is that we are blinded by our tendency to interpret artificial systems as human [51], [52]. This tendency makes it too easy to convince us of the intelligence, human-level or otherwise, of a machine. This is the *forensic problem of anthropomorphism* [28]. Anthropomorphizing risks introducing bias (in favour of the machine) into judgements of intelligence in machines – unless the risk is mitigated, these judgements are suspect.

III. ANTHROPOMORPHISM IN HRI

A. The scope of anthropomorphism

Considering the chronic tendency of people to anthropomorphize their environment, it should not be surprising that this topic gathered a lot of attention in the field of HRI. Reeves and Nass [53] in a series of experiments under the ‘Computers are Social Actors’ paradigm, showed that even computers are treated socially by people. In fact, robots seem to be especially well suited for benefiting from this phenomenon due to their higher anthropomorphisation compared to other technologies and physical autonomy in natural human environments [54], [55]. Collocated robots are anthropomorphized more than remote projected robots or embodied conversational agents [56], which emphasizes the importance of physical presence.

Robots’ embodiment should be always designed in a way that matches their tasks [57]. Both the embodiment and degrees of freedom can influence HRI. However, the former affects the degree to which a robot is perceived as an interaction partner, while the latter influences the way users perceive suitability of a robot for a current task [58]. Moreover, Hegel et al. [59] showed that people implicitly attribute human-like qualities to nonhuman agents. A robot’s embodiment affects the perception of their intelligence and intentionality on neurophysiological and behavioral levels. In addition the visual-cognition system allocates different level of visual attention depending on a robot’s embodiment [60]. Animate (anthropomorphic and zoomorphic) robots attract more attention than inanimate robots. Furthermore, a more human-like physical appearance of a robot can increase the empathy expressed by people towards it [61]. It is easier to relate to a robot that shares physical similarities with a human than with one that resembles a machine.

However, anthropomorphism defined in HRI as attribution of humanlike properties or characteristics to real or imagined nonhuman agents and objects [62] is not only affected by a robot’s physical appearance. Numerous other factors have been shown to affect perceived human-likeness of a robot, such as movement [63], verbal communication [64], [65], emotions

[66], gestures [67] and intelligence [68], [69]. Moreover, there are also user related factors determining the level of robot's anthropomorphism, such as motivation [70], social and cultural background [71], gender [72], group membership [73]. Finally, mere interaction with a robot can lead to higher anthropomorphization of it [74].

B. The importance of anthropomorphism in HRI

If anthropomorphism is such a widespread phenomenon and there are numerous factors affecting perceived humanlikeness of a robot, why would we pay so much attention to it? Why does HRI put such an effort to understand it better? Why does the HRI community not just accept the fact that anthropomorphism exists and focus their efforts on other areas? The potential answer could be deduced from the studies of anthropomorphism's impact on HRI. humanlikeness plays an important role in shaping the interaction. It can lead to decreased empathy for machine-like robots and can be responsible for harsher treatment of them compared to humanoids. Bartneck et al. [75] found that robots high on anthropomorphism/zoomorphism were praised more and punished less compared to machinelike robots, computer or human. In another study, a Sony's Aibo robot-dog was praised more than a human partner, but punished just as much [76]. Similarly, a machine-like robot was abused more than a human by receiving the highest voltage punishment [77].

In addition, human-like looking robots evoke reactions and expectations that are similar to human-human interaction. A humanoid is expected more than a mechanoid to adhere to human proxemic norms in HRI [78]. However, at least on a short-term, the violation of these norms can be counteracted by the higher reward-value of interacting with an anthropomorphic robot. Furthermore, the level of a robot's anthropomorphism can affect a patient's embarrassment during a medical checkup [79]. Finally, just the mere presence of a robot can lead to a social facilitation effect – a participant's better performance in easy and worse in difficult tasks [80]. Duffy [8] emphasized the importance of considerate design and use of robots' anthropomorphism in order to form meaningful interactions between people and robots. He proposed that anthropomorphism should not be used as a solution to all HRI problems, but rather a means that can facilitate the interaction when it is beneficial.

Anthropomorphism can also affect the acceptance of robots, which at the end can be a decisive factor for introducing social and service robots in natural human environments and their further development. This leads to an important question. What is the relationship between anthropomorphism and acceptance? The answer is rather complex and still not well understood. It is not enough to create an android and create robots that resemble humans. An undoubtedly single theory that received the most attention and tries to address this problem is the uncanny valley hypothesis.

IV. THE UNCANNY CURVE

The uncanny valley hypothesis [81] proposes a non-linear relationship between a robot's anthropomorphism and likeability. It suggests that making robots that look more humanlike will increase their likeability. However, when the gap between

a robot and human becomes really small the emotional reaction will instantly become strongly negative. Once the appearance and motion become indistinguishable from real humans the liking of a robot will be the same as for humans. Movement of a robot is expected to amplify the emotional response in comparison to static robots.

Although this theory gathered a lot of attention in science and mass media alike, there is relatively little empirical proof supporting it [82]. MacDorman [83] created a series of pictures that included a various degree of morph of a robot and a human. However, the outcome included images of beings that are not realistic and therefore it should not be surprising that they were found unfamiliar to participants. A potential explanation of the uncanny valley theory was provided by Saygin et al. [84], who found that compared to a human or machine-like robot, a mechanical movement of an android leads to higher activation of the human action perception system using fMRI. In other words, on the neurological level an android is not predictable as its mechanistic movement does not fit with human appearance.

However, the uncanny valley theory received also criticism in the recent years. Bartneck et al. [55] found that a highly realistic robot (android) is liked as much as a human. They concluded that anthropomorphism and likeability may be multi-dimensional constructs and therefore they cannot be projected on two dimensional space. Although, Ho and MacDorman [85] pointed out that the scales used in that research were correlated with *warmth* and as a result with each other.

In another study toy robots and humanoids were liked more than androids and humans [86]. Based on this finding the authors proposed that an uncanny valley is rather an uncanny cliff, where even the most humanlike robots are liked less than toy robots or mechanoids. That would imply that the attempts to build highly human-like androids might be be unfruitful, as they will have lesser chance of being accepted than robots with some mechanical features.

We believe that the proposed cliff is a result of having a higher probability of designing something eerie either in appearance, movement or interaction in an android than a more mechanistic robot. This relation between design complexity and likeability is presented in Figure 1. With greater humanlikeness there are more subtle ways to get it wrong and there can be something about these subtle discrepancies that is especially disconcerting. Potentially people are looking for features that distinguish androids from humans and even a slight difference may lead to their rejection. Considering the complexity of recreating a human being, there are multiple factors that can be done wrong. However, this problem does not exist for non-androids. They are easily distinguishable from humans and therefore they are not compared with them. However, any human-like features that they may have, make them more human-like and ultimately more liked. In other words, we believe that androids are compared with humans and humanoids and robotic toys likened to humans.

We perceive the following three topics as crucial for the research on the uncanny curve to progress: fixing terminology, finding the point at which a robot will start being compared with a human and its likeability will drop, and investigating

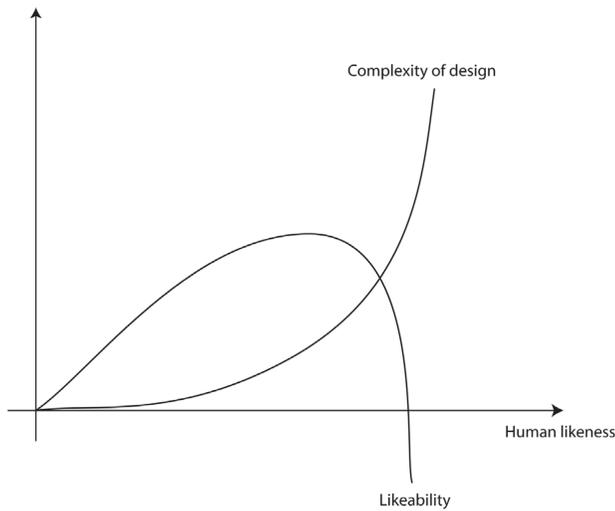


Fig. 1. The relation between complexity of design and likeability. In order to achieve higher level of human-likeness, the complexity of design must be increased to a level where possibility of doing something wrong is too high. The failure to achieve the design goals leads to decreased likeability of a robot.

the entities that lie between the deepest point of the uncanny valley and the human level.

The first issue raised by us is important as numerous terms were used for the y axis of the uncanny valley. The Japanese term used by Mori (*Shinwankan*) [81] is particularly difficult to translate to English. As a result human-likeness in the uncanny valley was related to familiarity [83], likeability [55], affinity [87], eeriness [85] or empathy [88]. Since different studies are using different terms it reduces their comparability and makes it harder to draw conclusions on this phenomenon. The discussion is still open regarding which term is the most appropriate. However, a commonly accepted conclusion would be more than welcome to avoid diversification in the work of different research teams as ultimately the efforts might be spent on a wrong goal. Rather than trying to fit data to a hypothetical graph and changing the terminology when it does not work, we should reverse the process and try to base our hypothesis on existing data. Moreover, it is possible that the term used by Mori is not the best one and therefore the issue is not just a problem of translation.

The second topic will help to understand the degree to which a robot should be made human-like for its optimal likeability. It can help us understand the concept of the uncanny curve by indicating what are the characteristics that make a robot too human-like for our liking. While the attempts to make a robot more anthropomorphic beyond this inflection

point might lead to its lower acceptance, we believe that it is still worth investigating creatures that are between a human and the most disliked forms of androids.

Up to now this sudden rise of likeability in the uncanny cliff has not been demonstrated. It is a terra incognita as the only comparison points in studies were humans. No entities were demonstrated to be similar enough to humans for their likeability to increase after passing the uncanny curve. Without that, the uncanny valley is merely an unproven hypothesis. We suggest that in order to explore that part of the spectrum it is necessary to understand the opposite process to anthropomorphization, known as dehumanization. (See also the psychologist Caporael's early work on 'mechanomorphism' – 'the attribution of characteristics of machines to humans' [89, p. 216].)

V. DEHUMANIZATION

The process of dehumanization – 'a failure to attribute basic human qualities to others' [90] – only recently became a focus of interest in the field of social psychology. Haslam [91] proposed a model of dehumanization that involves two distinct senses of humanness: characteristics that are uniquely human and those which form human nature. Denying the former attributes leads to perception of humans as animal-like, while denying the latter makes them object or automata like. Uniquely human (UH) characteristics are what separates humans from animals, such as intelligence, emotion recognition or self-control. On the other hand, features that are typical of or central to humans are referred to as human nature (HN) characteristics, such as primary emotions, warmth or personality. Therefore, characteristics that form the core of humanness may not be the same as those which distinguish us from other species.

There are several aspects that differentiate these two senses of humanness:

- UH characteristics reflect socialization and culture, while HN characteristics link humans to their inborn biological dispositions.
- UH characteristics reflect social learning and can vary across cultures and populations. HN is prevalent within populations and universal across different cultures.
- HN is essential, inherent and natural, while UH may not be perceived as essential.
- UH involves refinement, civility, morality, and higher cognition. HN involves cognitive flexibility, emotionality, vital agency, and warmth.

There are also different consequences of depriving humans of UH and HN characteristics [91]:

- A person deprived of UH traits is perceived as disgusting, while lack of HN traits implies indifference and lack of empathy.
- Lack of UH characteristics leads to perception of a person as subhuman and lack of HN as nonhuman.
- Deprivation of HN affects intergroup contexts and of UH affects intergroup and interpersonal contexts [92].

An important notion is that dehumanization does not only occur in extreme situation, but is rather common in its milder forms in our everyday social life [91]. Some social groups are implicitly and explicitly attributed less UH characteristics (e.g. artists) and therefore likened to animals, while other (e.g. businesspeople) are attributed less HN characteristics and likened to automata [93].

Understanding how dehumanization affects perception of humanness can give a new perspective on the process of anthropomorphization in HRI. It provides an indication of what characteristics may affect a user's perception of the degree to which a robot is perceived as human-like, and tools and methods used to measure dehumanization can be used in HRI for measuring anthropomorphism. The first publications appearing in HRI prove the potential of this approach [66], [67]. Another form of dehumanization can be also found in [86] where some of the images of human faces were modified with a slightly green hue to produce a mildly artificial look.

Moreover, we believe that using attributes of dehumanization can permit studies of the uncanny curve in actual HRI. Most of the studies conducted up to date suffered from at least one of the following problems:

- Images of robots were used rather than real robots.
- Static images or videos did not permit any interaction.
- Morphing images made unnatural creatures that cannot be found in the real world.
- The studies did not include long-term or even short-term HRI.

The first 3 problems lead to a potential lack of generalizability for HRI. It is possible that the uncanny valley, indicated by studies involving images, will have no relevance for an interaction between a user and robot. Furthermore, even if we assume that it impacts the interaction, its impact can be potentially marginal as the perceived human-likeness of a robot will change during the course of the interaction and therefore the effect will be limited to the very first few seconds of the interaction, e.g. Kiesler and Goetz [94] showed that robot's personality and speech can influence anthropomorphism more than embodiment. Moreover, the fourth problem indicated above is equally important. The perceived anthropomorphism of a robot is not a constant. It changes during the course of the interaction. Fussell et al. [74] showed that mere interaction with a robot leads to more anthropomorphic conceptions of robots. We notice the role that embodiment can have on a person's willingness to initiate an interaction with a robot. However, we believe that at least the same amount of effort, if not more, should be given to the interaction design rather than physical appearance design of a robotic platform, since the former's relation with the uncanny curve is even less known, despite the fact that the majority of interactions with robots are expected to last definitely longer than a few seconds.

To overcome these problems the studies should involve some form of HRI. The inclusion of attributes derived from studies of dehumanization can allow us to manipulate the level of anthropomorphism necessary to create entities that minimally differ from human beings in actual HRI. Therefore, the real test for the uncanny curve would be to see if it applies to robots behaving in human-like ways.

VI. CONCLUSIONS

We know that people anthropomorphize and have theories why it is such a common phenomenon. However, our understanding of its impact on HRI is in its infancy. Even the most-discussed hypothesis related to anthropomorphism in recent years, the uncanny valley, cannot be considered as anything more than an unproven theory. In this paper we pointed out some of the mistakes from which studies of the uncanny curve suffered:

- The lack of fixed terminology for y axis leads to poor comparability of studies.
- The inclusion of only images or videos of robots reduces the generalizability of the results for actual HRI.
- Considering anthropomorphism as something constant that does not change during the course of interaction has a lot of existing evidence of the opposite process to overcome.

Whereas all these issues on their own indicate that Mori's theory might receive more attention than it really should, there is a much bigger concern that should be addressed. The uncanny valley assumes that at some point entities that are sufficiently human-like will lead to increased likeability. However, to date this part of the valley has never been explored. This could be a statement about the power of media – everybody believes the curve to be the valley just because everybody else talks about it. However, there is no empirical evidence supporting the hypothesis. Therefore, rather than trying to fit data to the graph, we believe that the community should fit the graph to the available data. That is clearly an uncanny curve not valley.

Furthermore, the left side of the uncanny curve is rather well explored. The really interesting part of the graph is the right side as it has never been shown to exist. What we propose is to use the attributes from studies of dehumanization in order to affect the perceived human-likeness of robots in the real HRI rather than static images. These robots might be able to possess enough human-like qualities to pass the uncanny bottom in the interaction.

Finally, the question is whether there can be entities beyond the human on the graph. Are biological humans the ultimate end of humanness or can there be entities more 'human' than biological humans? Several influential researchers in AI currently forecast 'software-based humans'; they promise immortality for all, in a virtual state or implemented in a cybernetic body. Will software-based humans be *human*? And even if these forecasts are fantasy, it is certainly possible that future humans will have invisible implants that provide superpowers, such as extreme strength enabling a person to lift a car with one hand. Where should we place such a person on the uncanny curve?

REFERENCES

- [1] J. H. Lesher *et al.*, *Xenophanes of Colophon: Fragments: A Text and Translation With a Commentary*. University of Toronto Press, 2001, vol. 4.
- [2] S. Guthrie, *Faces in the clouds: A new theory of religion*. Oxford University Press, USA, 1995.

- [3] W. Brennan, *Dehumanizing the vulnerable: when word games take lives*, ser. Values & ethics series. Loyola University Press, 1995.
- [4] L. Abend, "In Spain, human rights for apes," *TIME.com*. [www.time.com/time/world/article/0, vol. 8599, no. 1824206, 2008](http://www.time.com/time/world/article/0,vol.8599,no.1824206,2008).
- [5] E. O. Wilson, *The creation: An appeal to save life on earth*. Wiley Online Library, 2006.
- [6] J. Aaker, "Dimensions of brand personality," *Journal of Marketing research*, pp. 347–356, 1997.
- [7] P. Aggarwal and A. McGill, "Is that car smiling at me? schema congruity as a basis for evaluating anthropomorphized products," *Journal of Consumer Research*, vol. 34, no. 4, pp. 468–479, 2007.
- [8] B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 177–190, 2003.
- [9] C. Darwin, "1998. the expression of the emotions in man and animals," 1872.
- [10] R. Hard, *The Routledge Handbook of Greek Mythology: Based on HJ Rose's "Handbook of Greek Mythology"*. Psychology Press, 2004.
- [11] F. Heider and M. Simmel, "An experimental study of apparent behavior," *The American Journal of Psychology*, vol. 57, no. 2, pp. 243–259, 1944.
- [12] D. Hume, *The Natural History of Religion (1757)*, H. Root, Ed. London: Adam and Charles Black, 1956.
- [13] J. Bering, *Origins of the Social Mind: Evolutionary Psychology and Child Development*. New York: The Guildford Press, 2005, ch. The Evolutionary History of an Illusion, pp. 411–437.
- [14] J. M. Bering, "The folk psychology of souls," *Behavioral and Brain Sciences*, vol. 29, pp. 453–462, 2006.
- [15] J. Bering, *The God Instinct*. London: Nicholas Brealey, 2010.
- [16] J. Barrett, *Why Would Anyone Believe in God?* Lanham, MD: AltaMira Press, 2004.
- [17] P. Boyer, *Religion Explained*. New York: Basic Books, 2001.
- [18] A. Landtblom, "The sensed presence: an epileptic aura with religious overtones," *Epilepsy & Behavior*, vol. 9, no. 1, pp. 186–188, 2006.
- [19] S. Arzy, M. Seeck, S. Ortigue, L. Spinelli, and O. Blanke, "Induction of an illusory shadow person," *Nature*, vol. 443, no. 7109, pp. 287–287, 2006.
- [20] E. Poulet, J. Brunelin, B. Bediou, R. Bation, L. Forgeard, J. Dalery, T. D'Amato, and M. Saoud, "Slow transcranial magnetic stimulation can rapidly reduce resistant auditory hallucinations in schizophrenia," *Biological psychiatry*, vol. 57, no. 2, pp. 188–191, 2005.
- [21] J. L. Saver and J. Rabin, "The neural substrates of religious experience," *Journal of Neuropsychiatry and Clinical Neurosciences*, vol. 9, no. 3, pp. 498–510, 1997.
- [22] M. Trimble and A. Freeman, "An investigation of religiosity and the gastaut-geschwind syndrome in patients with temporal lobe epilepsy," *Epilepsy & behavior*, vol. 9, no. 3, pp. 407–414, 2006.
- [23] P. Granqvist, M. Fredrikson, P. Unge, A. Hagenfeldt, S. Valind, D. Larhammar, and M. Larsson, "Sensed presence and mystical experiences are predicted by suggestibility, not by the application of transcranial weak complex magnetic fields," *Neuroscience letters*, vol. 379, no. 1, pp. 1–6, 2005.
- [24] J. L. Barrett, "Exploring the natural foundations of religion," *Trends in cognitive sciences*, vol. 4, no. 1, pp. 29–34, 2000.
- [25] J. Barrett, "Cognitive science of religion: What is it and why is it?" *Religion Compass*, vol. 1, no. 6, pp. 768–786, 2007.
- [26] P. Boyer, "Religious thought and behaviour as by-products of brain function," *Trends in cognitive sciences*, vol. 7, no. 3, pp. 119–124, 2003.
- [27] A. Turing, *The Essential Turing*. Oxford: Oxford University Press, 2004, ch. Intelligent Machinery (1948).
- [28] D. Proudfoot, "Anthropomorphism and ai: Turing's much misunderstood imitation game," *Artificial Intelligence*, vol. 175, no. 5, pp. 950–957, 2011.
- [29] —, "Rethinking turing's test," *Journal of Philosophy*, To appear.
- [30] B. Scassellati, *How social robots will help us to diagnose, treat, and understand autism*, ser. Springer Tracts in Advanced Robotics, 2007, vol. 28.
- [31] J. Fasola and M. J. Matarić, "Using socially assistive human-robot interaction to motivate physical exercise for older adults," vol. 100, no. 8, Piscataway, NJ, United States, 2012, pp. 2512 – 2526.
- [32] E. Wade, A. R. Parnandi, and M. J. Matarić, "Using socially assistive robotics to augment motor task performance in individuals post-stroke," in *IEEE International Conference on Intelligent Robots and Systems*, 2011, pp. 2403–2408.
- [33] D. Feil-Seifer and M. J. Matarić, "Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features," in *HRI 2011 - Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, 2011, pp. 323–330.
- [34] N. Giullian, D. Ricks, A. Atherton, M. Colton, M. Goodrich, and B. Brinton, "Detailed requirements for robots in autism therapy," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 2595–2602.
- [35] C. Breazeal, "Human-robot partnership," *IEEE Intelligent Systems*, vol. 21, no. 4, pp. 79–81, 2006.
- [36] A. Tapus, M. J. Matarić, and B. Scassellati, "Socially assistive robotics [grand challenges of robotics]," *IEEE Robotics and Automation Magazine*, vol. 14, no. 1, pp. 35–42, 2007.
- [37] J. Bates, "Role of emotion in believable agents," *Communications of the ACM*, vol. 37, no. 7, pp. 122–125, 1994.
- [38] A. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.
- [39] R. Descartes, *The Philosophical Writings of Descartes: Volume 1*. Cambridge: Cambridge University Press, 1985, ch. Discourse on the Method (1637).
- [40] C. Breazeal, "Early experiments using motivations to regulate human-robot interaction," in *AAAI Fall Symposium on Emotional and Intelligent: The tangled knot of cognition, Technical Report FS-98-03*, 1998, pp. 31–36.
- [41] C. Breazeal and P. Fitzpatrick, "That certain look: Social amplification of animate vision," in *Proceedings of the AAAI Fall Symposium on Society of Intelligence Agents: The Human in the Loop*, 2000.
- [42] D. Proudfoot, *The Role and Use of Conceptual Analysis in Psychology: A Wittgensteinian Perspective*. New York: Palgrave Macmillan, To appear, ch. Can a Robot Smile? Wittgenstein on Facial Expression.
- [43] B. Scassellati, "How robotics and developmental psychology complement each other," in *NSF/DARPA Workshop on Development and Learning*, 2000.
- [44] —, "Theory of mind for a humanoid robot," *Autonomous Robots*, vol. 12, no. 1, pp. 13–24, 2002.
- [45] K. Gold and B. Scassellati, "A bayesian robot that distinguishes self from other," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, 2007.
- [46] C. Breazeal, J. Gray, and M. Berlin, "An embodied cognition approach to mindreading skills for socially intelligent robots," *International Journal of Robotics Research*, vol. 28, no. 5, pp. 656–680, 2009.
- [47] C. Breazeal, "Role of expressive behaviour for robots that learn from people," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3527–3538, 2009.
- [48] D. Proudfoot and B. J. Copeland, *Oxford Handbook of Philosophy of Cognitive Science*. New York: Oxford University Press, 2012, ch. Artificial Intelligence, pp. 147–82.
- [49] J. B. Pollack, "Mindless intelligence," *IEEE Intelligent Systems*, vol. 21, no. 3, pp. 50–56, 2006.
- [50] K. Ford and P. Hayes, "On computational wings: Rethinking the goals of artificial intelligence—the gold standard of traditional artificial intelligence—passing the so-called turing test and thereby appearing to be," *Scientific American Presents*, vol. 9, no. 4, p. 79, 1998.
- [51] D. Proudfoot, "How human can they get?" *Science*, vol. 284, no. 5415, pp. 745–745, 1999.
- [52] —, *Alan Turing: Life and Legacy of a Great Thinker*. Berlin: Springer-Verlag, 2004, ch. Robots and Rule-following, pp. 359–79.
- [53] B. Reeves and C. Nass, "The media equation," 1996.
- [54] S. Kiesler and P. Hinds, "Introduction to this special issue on human-robot interaction," *Hum.-Comput. Interact.*, vol. 19, no. 1, pp. 1–8, Jun. 2004.

- [55] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita, "My robotic doppelgänger - a critical look at the uncanny valley," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2009, pp. 269–276.
- [56] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, "Anthropomorphic interactions with a robot and robot-like agent," *Social Cognition*, vol. 26, no. 2, pp. 169–181, 2008.
- [57] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, oct.-2 nov. 2003, pp. 55 – 60.
- [58] K. Fischer, K. S. Lohan, and K. Foth, "Levels of embodiment: Linguistic analyses of factors influencing hri," in *HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2012, pp. 463–470.
- [59] F. Hegel, S. Krach, T. Kircher, B. Wrede, and G. Sagerer, "Understanding social robots: A user study on anthropomorphism," in *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, 2008, pp. 574–579.
- [60] J. Bae and M. Kim, "Selective visual attention occurred in change detection derived by animacy of robot's appearance," in *Proceedings of the 2011 International Conference on Collaboration Technologies and Systems, CTS 2011*, 2011, pp. 190–193.
- [61] L. D. Riek, T. Rabinowitch, B. Chakrabarti, and P. Robinson, "How anthropomorphism affects empathy toward robots," in *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, HRI'09*, 2009, pp. 245–246.
- [62] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: A three-factor theory of anthropomorphism," *Psychological review*, vol. 114, no. 4, pp. 864–886, 2007.
- [63] E. Wang, C. Lignos, A. Vatsal, and B. Scassellati, "Effects of head movement on perceptions of humanoid robot behavior," in *HRI 2006: Proceedings of the 2006 ACM Conference on Human-Robot Interaction*, 2006, pp. 180–185.
- [64] V. K. Sims, M. G. Chin, H. C. Lum, L. Upham-Ellis, T. Ballion, and N. C. Lagattuta, "Robots' auditory cues are subject to anthropomorphism," in *Proceedings of the Human Factors and Ergonomics Society*, vol. 3, 2009, pp. 1418–1421.
- [65] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, and R. Te Boekhorst, "Human approach distances to a mechanical-looking robot with different robot voice styles," in *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, 2008, pp. 707–712.
- [66] F. Eyssel, F. Hegel, G. Horstmann, and C. Wagner, "Anthropomorphic inferences from emotional nonverbal cues: A case study," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2010, pp. 646–651.
- [67] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joubin, *Effects of gesture on the perception of psychological anthropomorphism: A case study with a humanoid robot*, ser. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2011, vol. 7072 LNAI.
- [68] C. Bartneck, M. Verbunt, O. Mubin, and A. Al Mahmud, "To kill a mockingbird robot," Arlington, VA, United states, 2007, pp. 81 – 87.
- [69] C. Bartneck and J. Hue, "Exploring the abuse of robots," *Interaction Studies*, vol. 9, no. 3, pp. 415–433, 2008.
- [70] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: A three-factor theory of anthropomorphism," *Psychological review*, vol. 114, no. 4, pp. 864–886, 2007.
- [71] V. Evers, H. C. Maldonado, T. L. Brodecki, and P. J. Hinds, "Relational vs. group self-construal: Untangling the role of national culture in hri," in *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, 2008, pp. 255–262.
- [72] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. De Ruyter, and F. Hegel, "'if you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism," in *HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2012, pp. 125–126.
- [73] D. Kuchenbrandt, F. Eyssel, S. Bobinger, and M. Neufeld, "Minimal group-maximal effect? evaluation and anthropomorphization of the humanoid robot nao," *Social Robotics*, pp. 104–113, 2011.
- [74] S. R. Fussell, S. Kiesler, L. D. Setlock, and V. Yew, "How people anthropomorphize robots," in *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, 2008, pp. 145–152.
- [75] C. Bartneck, J. Reichenbach, and J. Carpenter, "Use of praise and punishment in human-robot collaborative teams," Piscataway, NJ, USA, 2006.
- [76] J. Reichenbach, C. Bartneck, and J. Carpenter, "Well done, robot! - the importance of praise and presence in human-robot collaboration," Hatfield, United kingdom, 2006, pp. 86–90.
- [77] C. Bartneck, C. Rosalia, R. Menges, and I. Deckers, "Robot abuse—a limitation of the media equation," in *Proceedings of the Interact 2005 Workshop on Agent Abuse, Rome*, 2005.
- [78] D. S. Syrdal, K. Dautenhahn, M. L. Walters, and K. L. Koay, "Sharing spaces with robots in a home scenario - anthropomorphic attributions and their effect on proxemic expectations and evaluations in a live hri trial," in *AAAI Fall Symposium - Technical Report*, vol. FS-08-02, 2008, pp. 116–123.
- [79] C. Bartneck, T. Bleeker, J. Bun, P. Fens, and L. Riet, "The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots," *Paladyn*, pp. 1–7, 2010.
- [80] N. Riether, F. Hegel, B. Wrede, and G. Horstmann, "Social facilitation with social robots?" in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: ACM, 2012, pp. 41–48.
- [81] M. Mori, "The uncanny valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.
- [82] K. Ogawa, C. Bartneck, D. Sakamoto, T. Kanda, T. Ono, and H. Ishiguro, "Can an android persuade you?" Piscataway, NJ, USA, 2009, pp. 516–21.
- [83] K. MacDorman, "Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it," in *CogSci-2005 workshop: toward social mechanisms of android science*, 2005, pp. 106–118.
- [84] A. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith, "The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions," *Social cognitive and affective neuroscience*, vol. 7, no. 4, pp. 413–422, 2012.
- [85] C. Ho and K. MacDorman, "Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1508–1518, 2010.
- [86] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita, "Is the uncanny valley an uncanny cliff?" Piscataway, NJ, USA, 2007, pp. 368–73.
- [87] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley," *IEEE Robotics and Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.
- [88] C. Misselhorn, "Empathy with inanimate objects and the uncanny valley," *Minds and Machines*, vol. 19, no. 3, pp. 345–359, 2009.
- [89] L. R. Caporael, "Anthropomorphism and mechanomorphism: Two faces of the human machine," *Computers in Human Behavior*, vol. 2, no. 3, pp. 215–234, 1986.
- [90] A. Waytz and N. Epley, "Social connection enables dehumanization," *Journal of experimental social psychology*, vol. 48, no. 1, pp. 70–76, 2012.
- [91] N. Haslam, "Dehumanization: An integrative review," *Personality and Social Psychology Review*, vol. 10, no. 3, pp. 252–264, 2006.
- [92] N. Haslam, P. Bain, L. Douge, M. Lee, and B. Bastian, "More human than you: Attributing humanness to self and others," *Journal of personality and social psychology*, vol. 89, no. 6, pp. 937–950, 2005.
- [93] S. Loughnan and N. Haslam, "Animals and androids: Implicit associations between social categories and nonhumans," *Psychological Science*, vol. 18, no. 2, pp. 116–121, 2007.
- [94] S. Kiesler and J. Goetz, "Mental Models and Cooperation with Robotic Assistants," in *Proc. of Conference on Human Factors in Computing Systems*, 2002, pp. 576–577.