Between Anthropomorphism, Trust, and the Uncanny Valley: A Dual-Processing Perspective on Perceived Trustworthiness and Its Mediating Effects on Use Intentions of Social Robots

Anika Nissen
University Duisburg-Essen
anika.nissen@uni-due.de

Katharina Jahn
University of Siegen
katharina.jahn@uni-siegen.de

Abstract

Designing social robots with the aim to increase their acceptance is crucial for the success of their implementation. However, even though increasing anthropomorphism is often seen as a promising way to achieve this goal, the uncanny valley effect proposes that anthropomorphism can be detrimental to acceptance unless robots are almost indistinguishable from humans. Against this background, we use a dual processing theory approach to investigate whether an uncanny valley of perceived trustworthiness (PT) can be observed for social robots and how this effect differs between the intuitive and deliberate reasoning system. The results of an experiment with four conditions and 227 participants provide support for the uncanny valley effect. Furthermore, mediation analyses suggested that use intention decreases through both reduced intuitive and deliberate PT for medium levels of anthropomorphism. However, for high levels of anthropomorphism (indistinguishable from real human), only intuitive PT determined use intention. Consequently, our results indicate both advantages and pitfalls of anthropomorphic design.

1. Introduction

In almost all situations of our lives, first impressions are made in the blink of an eye [1, 2] and often already predict our further attitude and behavior. The reason for this can be found in first impressions, especially of visual beauty, leading to a halo effect due to which further assumptions about the trustworthiness, warmth, and competence of a robot are made [1, 2]. Further, as social robots are designed increasingly similar to actual humans, anthropomorphism has shown to significantly correlate with evaluations of perceived trustworthiness (PT) [3, 4], a crucial predictor for use intentions [5, 6]. Moreover, positive affect towards robots, such as warmth and PT, are pivotal for humans to accept and adopt social robots in their life [7], which is a necessary step to enable comfortable, social human-robot interactions. Especially for the interaction with robots, adding emotional and social interactions tends to reduce the perceived stress and thus, increase PT in the robot with which the interaction took place [8]. However, while these aspects are also factors for increased human-likeness of robots, the correlation between anthropomorphism and positive affect towards robots does not follow a linear line but enters at a specific level an uncanny valley [9]. In the uncanny valley, human actors have increased negative attitudes towards robots, which become positive again when anthropomorphism is almost indistinguishable from a real human [9]. Even though the first introduction of the uncanny valley effect happened a century ago and the levels of anthropomorphism in robots have significantly increased since then (see i.e. the robot Erica), the uncanny valley effect seems to still hold true for higher anthropomorphic robots [10, 11].

While increased anthropomorphism has several positive effects, it might also facilitate humans to apply social reasoning towards robots (such as theory of mind). As a consequence, human users may cease to distinguish between humans and robots even though it would be necessary [12]. In this respect, the differentiation between two systems of social processing is crucial: (1) an intuitive, affective system, and (2) a cognitive, reflective system [13, 14]. Given the increasing levels of anthropomorphism in robots, the intuitive system might not be able anymore to make a distinction between human and robot, while i.e. the cognitive system might then detect the processing error [15, 16]. Given that PT is a complex construct which also consists of emotional and cognitive reasoning [17, 5], the dual-processing theory might therefore be transferable when considering PT as pivotal impact factor for social human-robot interactions [18]. However, to the best of our knowledge, this perspective has not yet been taken to investigate the mediating effects of anthropomorphism level, intuitive and deliberate PT on further use intentions.

URI: https://hdl.handle.net/10125/70654
978-0-9981331-4-0

Consequently, this paper aims to investigate, a) *if and how the uncanny valley of PT for human-like robots differs between intuitive and deliberate PT* and b) *how these PT types influence the use intentions of the robot*. To address this research goal, we first review literature regarding PT as a construct and how it can be understood in correspondence to a dual-processing theory. After that, we focus on PT specifically in human-robot interactions by taking robot anthropomorphism as one major influence factor on pre-interaction PT evaluations. Next, we design a study focusing on intuitive and deliberate pre-interaction PT evaluations of social robots on four different levels of anthropomorphism and investigate their mediating effects on use in form of interaction intentions. From the results of this study, we are able to sketch uncanny valleys for the intuitive and deliberate evaluations each, and identify the impact of both reasoning systems.

2. Related Literature

The construction and introduction of social robots receives more and more attention in various application fields such as education, support for decision makers, healthcare, therapy, at the workplace, or at home [19, 20, 21, 22]. In all of these application fields, the cooperation and collaboration with social robots is crucial for their successful implementation. Therefore, when designing robots for social interactions, there are several (unwritten) rules and norms which robots need to adhere to when they should be accepted in everyday life [23, 21]. For instance, robots require human spatial skills for moving naturally between humans [24], or they need the ability to recognize and express emotions and empathy [25]. While the degree of fulfilling social requirements is also severely influenced by how the robot behaves, its visual appearance already gives first cues which lead to expectations about its behavior [19]. That is, solely depending on the first impression of a robot's visual appearance, assumptions about its capabilities and roles in social contexts are automatically made [26, 27, 20]. However, designing for high visual anthropomorphism might not always have the desired effect, since it might lead to expectations about the robot which cannot be met during interaction [28]. Therefore, a focused investigation of the especially relevant construct of PT for high anthropomorphic robots seems reasonable.

2.1. Dual-Processing Theory and Perceived Trustworthiness

In social interactions, trust and PT are crucial and complex constructs which can be subdivided into different types. One approach from Information Systems research is the distinction between (1) trust beliefs, (2) trusting intention, and (3) disposition to trust [17, 5]. In this paper, we further focus on (1) trusting beliefs which are elicited by a robot's visual design features (in means of anthropomorphism) and which are in this paper referred to as PT. PT can be further subdivided i.e. into emotional vs. cognitive PT [18], which might also be integrated with one another [29].

Since PT is a social construct, this integration might be similar to the dual-processing theory of social reasoning, in which system 1 is characterized as affective, automatic, and intuitive processing, while system 2 is thought to be cognitive, rational, and deliberate processing [13]. Consequently, the often conceptualized emotional and cognitive component of PT might also be seen as system 1 and system 2 reasoning [18]. System 1, or in this paper further referred to as "intuitive PT" evaluates stimuli fast and is more prone to erroneous decisions than system 2 [30]. The processing in system 2, further referred to as "deliberate PT" can either support or contradict to what has already been evaluated in system 1 [13]; however, it can not completely inhibit system 1 [14].

By taking this dual-processing perspective, we make the implicit assumption that human-robot interactions require social reasoning and cognition. Apparently, while robots become more human-like, there is the belief that the social processing of robots will also become closer to that of a human [13]. Research stating that humans do not differentiate much between robots and humans when both are perceived as trustworthy [31] further supports this assumption. However, when it comes to how PT is evaluated, there are general differences between how we evaluate the PT of a human and how that of a technology [32]. This further implies the question, with which criteria the PT of a robot might be evaluated. Do users assess criteria closer to human-related factors, such as competence or benevolence, or are they more concerned about technological factors, such as reliability and helpfulness? Especially in relation to different levels of anthropomorphism and a possible uncanny valley, these questions might further help to gain deeper insights into how humans perceive and evaluate social robots. Thus, we take these factors into account for measuring deliberate PT, but will not discuss them further.

2.2. Anthropomorphism and the Uncanny Valley

Studies investigating the antecedents of human PT in robots have identified that both human characteristics

and ability, as well as robotic attributes and performance impact how trustworthy a robot is perceived [3, 7]. For instance, the personality traits of humans can significantly impact their attitudes towards the robot [33]. Among robot characteristics or attributes, anthropomorphism has also shown to significantly correlate with PT ratings [3] and generally can be represented through visual cues, auditory, or behavioral characteristics of robots [34]. In the frame of this work, however, we focus on the visual cues only, as this is about the first cue we perceive and evaluate from a robot, which may lead to starting a conversation and subsequently evaluating its speech and behavior.

When talking about anthropomorphism of robots, however, the uncanny valley effect also needs to be addressed [9]. This effect is one of the most relevant approaches to explain how individuals differentiate between humans and robots, or more precisely, between different levels of anthropomorphism [9, 6]. This effect proposes that if anthropomorphism reaches a certain level between high and low human-likeness, the *uncanny valley* is entered [35], which results in more negative reactions against this entity. Only when anthropomorphism is high and becomes almost indistinguishable from a human, the impression becomes more positive again. While changes in the evaluation of robots between first impression and first interaction are likely [36], we will further focus on first impression and its uncanny valley, only. Further, since Marthur et al. [37] have already shown for a variety of robot faces that the uncanny valley effect generally holds true when affect and PT are evaluated in first impressions, we further explicitly consider PT as dual-process and investigate the differences between intuitive and deliberate PT evaluations not only in regard to an uncanny valley, but also regarding their mediating effects on use/interaction intentions.

3. Method

3.1. Sample

We used the online platform clickworker to recruit participants for the survey. 253 participants completed the questionnaire. After excluding participants who were faster than 90% of the sample, to remove participants who merely clicked through the questionnaire without answering the questions seriously, 227 participants remained. 60.79% of participants were male and 38.76% were female and one participant was non-binary/diverse (0.44%). Additionally, participants were between 18 and 69 years old ($M = 37.83$, $SD = 11.74$). The majority of participants indicated to be working in the services sector (17.62%), followed by the IT sector (13.22%) and business administration sector (9.69%) The remaining participants (59.47%) worked in a diverse set of other sectors or were currently unemployed.

3.2. Stimuli

In order to investigate possible uncanny valleys existing for intuitive and deliberate PT, we included robots at four different anthropomorphism levels in our study. Since women tend to be perceived as being more trustworthy [38], we decided to include mainly female robots. More precisely, we included Nao as representative control group for low anthropomorphism which is at the same time the only gender-less robot condition. For higher degrees of anthropomorphism, we used Sophia for medium-low anthropomorphism, Mark1 for medium-high anthropomorphism, and finally, a human for the high anthropomorphism (Figure 1). At this point, it needs to be noted that participants were told that all photographs illustrated robots, including the human.

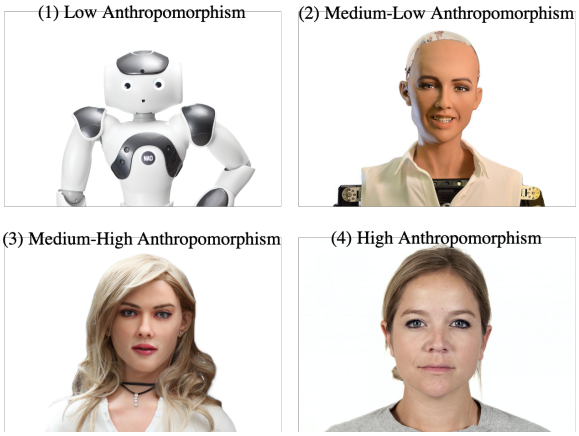

Figure 1. Stimuli Used in the Survey

3.3. Study Design

To investigate how intuitive and deliberate PT are evaluated based on a robot's visual appearance, we created an online survey designed as follows. First, participants had to rate their general attitude towards technology and their PT perceptions regarding technology, innovations, and humans. After that, pictures of the robots were shown and intuitive PT was measured with time restrictions for the decision to make. This was followed by demographic questions, in which a control question was included to filter out inattentive participants. In case the control question was answered

	Trust Intention	Integrity	Benevolence	Competence	Reliability	Helpfulness
Integrity	0.79					
Benevolence	0.80	0.83				
Competence	0.76	0.79	0.79			
Reliability	0.78	0.82	0.75	0.80		
Helpfulness	0.80	0.83	0.84	0.90	0.84	
Functionality	0.80	0.82	0.82	0.88	0.82	0.90

Table 1. Correlations between PT Scales

wrong, the questionnaire was closed and participants could not continue. If the control question was answered correctly, participants proceeded to rate the robot conditions for deliberate PT for which the pictorial stimuli were shown and had to imagine they had the opportunity to use the robot in a shopping situation before the scales appeared. Finally, participants answered the humanness manipulation check, after which they were thanked, debriefed and received their clickworker code. This study design is justified by the visual robot appearance having a significant impact on trust evaluations, which is already formed at the first impression of the robot and significantly impacts further evaluations of the robot [1, 2]. Therefore, we use images of robots as stimuli which provide us a first indication of how PT of the robots is evaluated.

3.4. Measurements

We included a single item scale for intuitive PT that had to be rated on a 5-point Likert scale ("To which degree would you trust this robot", from 5=completely to 1=not at all). To avoid biases due to the single item, we repeated this question three times for each included robot (i.e., participants completed 12 trials in total) in randomized order and averaged the answers across the three trials for each robot for further analyses. Deliberate PT was measured primarily with the three items for trusting intentions scale from McKnight et al. [39] used by [32] (e.g., "I can always rely on this robot for buying new products"). Since this construct might not fully capture deliberate PT, we additionally included the PT scales [32] of human-PT being integrity, competence, and benevolence [39] and of technology-PT being functionality, helpfulness, and reliability [40]. All of these scales, however, were only shown for one randomly selected robot (leading to 53 data points for low, 56 for medium-low, and 59 each for medium-high and high condition), while the perceived trusting intentions scale was shown for every robot so that the questionnaire did not get too tiring. Finally, we included a use intention scale for every robot, which was adapted from Davis et al. [41] (e.g., "I would use this robot to assist me in my buying decision") to further interpret how intuitive and deliberate PT influence use intentions and thus, further HRI. Cronbach's Alpha indicated sufficient reliability for all three scales (intuitive PT: .87, between .94 and .97 for individual robots; deliberate PT: .94, between .91 to .93 for individual robots; use intention: .95, between .93 and .94 for individual robots) [42]. The manipulation check for humanness ("Please indicate to which degree the robots pictured below look like a machine or a person to you") consisted of rating the humanness of each stimuli on a scale from 0% (= machine-like) to 100% (= person-like).

To ensure that intuitive and deliberate processing were primarily measured, we used two means: (1) a time pressure/time delay component and (2) one-item question for intuitive trustworthiness, and a multi-item questionnaire for the deliberate trustworthiness. Using a time component in a questionnaire to distinguish between intuitive and reflective system has shown to be a common method in other studies [43, 44, 30]. Therefore, image and scale for the intuitive PT scale was shown for merely 4 seconds in which participants had to make a decision. In case they did not make an input within the 4 seconds, the system remained at the question and robot. For the deliberate PT, first, the image of the robot was shown alone and after 5 seconds the scales appeared in addition to the image and were then clickable. Through this, we tried to ensure that participants looked at the stimuli for a certain time before evaluating the deliberate PT scales taken from McKnight et al. [39, 40].

4. Results

Means and standard deviations are given in Table 2. Because the assumption of sphericity was violated for intuitive and deliberate PT, we analyzed the data using oneway repeated measures ANOVAs with Greenhouse-Geisser correction. Additionally, we used Tukey-corrected post-hoc tests for follow-up analyses.

Because we measured only PT intention in a within design, we chose to use it as our main outcome for overall PT. To validate that it covers different facets of PT, we looked at bivariate correlations. All PT scales

Anthropomorphism	Low M (SD)	Medium-Low M (SD)	Medium-High M (SD)	High/Human M (SD)
Intuitive PT	2.94 (1.08)	2.27 (0.91)	2.53 (1.03)	3.70 (1.03)
Deliberate PT	3.86 (1.51)	3.13 (1.47)	3.34 (1.51)	4.30 (1.53)
Use Intention	4.02 (1.60)	3.25 (1.59)	3.51 (1.54)	4.48 (1.56)

Table 2. Means and Standard Deviations for Intuitive and Deliberate PT

were significantly correlated, with a minimum value of .75 and the minimum value for PT intention was at .76. Therefore, the measures were highly correlated which lets us assume that the trust intention scale sufficiently represents the different PT facets as a measure.

4.1. Manipulation Checks

The ANOVA for the manipulation check of anthropomorphism was significant ($F(2.52, 559.74) = 1106.98, p < .001, \eta_G = .735$). Post-hoc tests showed that the high condition was seen as most human-like (94.97%), followed by the medium-high condition (54.15%), the medium-low condition (30.43%), and then the low condition (7.62%) (all with $p < .001$, on a scale from 0% being machine-like and 100% being person-like). Therefore, the anthropomorphism manipulation was successful.

4.2. Experimental Results

Intuitive PT: The repeated measures ANOVA revealed a significant main effect (F (2.64, 597.27) = 123.56, $p < .001$) pointing to an uncanny valley effect. That is, post-hoc tests showed that this was due to higher PT ratings for the high condition than for any other condition, followed by the low anthropomorphism condition, the medium-high anthropomorphism condition, and, finally, the medium-low condition. All ps were $< .001$ except for the difference between the medium-low and medium-high conditions, which was at .007. Consequently, these results support the uncanny valley effect for the intuitive PT for which current high anthropomorphism levels still do not seem to be high enough.

Deliberate PT: For deliberate PT, a similar pattern emerged. Post-hoc tests after the significant ANOVA ((F (2.73, 616.99) = 63.95, $p < .001$) showed that the high anthropomorphism condition had again the highest PT ratings, followed by the low anthropomorphism condition (all $ps < .001$). However, in contrast to intuitive PT, no significant difference could be found between the medium-low and medium-high condition ($p = .120$). These results support the uncanny valley effect also exists for deliberate PT ratings, albeit rated generally higher than intuitive PT. A graphical overview on the detected uncanny valley effects is given in Figure 2 which shows that both valleys seem to be mostly parallel for the anthropomorphism levels. Further, since PT is a crucial impact factor for robot acceptance and further use intentions, we investigate the impact of anthropomorphism levels, intuitive PT, and deliberate PT on use intention in the following model (Figure 3).

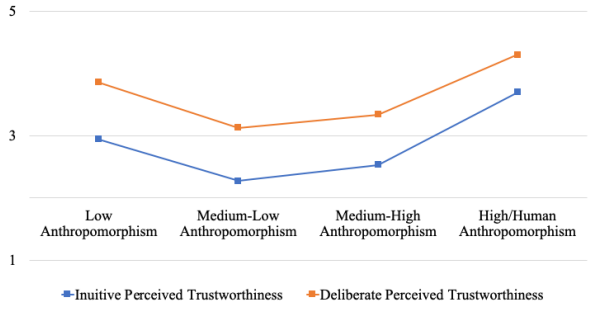

Figure 2. Uncanny Valleys of Intuitive and Deliberate PT

4.3. Mediation Analysis

We used multilevel mediation analyses in a 1-1-1 mediation [45] with the R package mediation [46] and lme4 to check to which degree deliberate and intuitive PT contribute to explain the effects of the robots' anthropomorphism on reuse intention. The intraclass correlation (ICC) indicated that using a multilevel approach is necessary for both deliberate PT (ICC = .49) and intuitive PT (ICC = .15). The anthropomorphism conditions were dummy coded, with the low condition coded as 0 and the other three robots coded as 1. We used a model with random intercepts and fixed slopes. The overall results of the mediation model are displayed in Figure 3.

Deliberate PT. For deliberate PT, a multilevel model with PT as dependent variable and the anthropomorphism conditions as well as intuitive PT as predictors revealed that the medium-low condition ($beta = -.17$, $p = .001$) and the medium-high condition ($beta = -.15$, $p < .003$) were perceived less trustworthy than the low condition, whereas the high condition did not show a significant difference from the

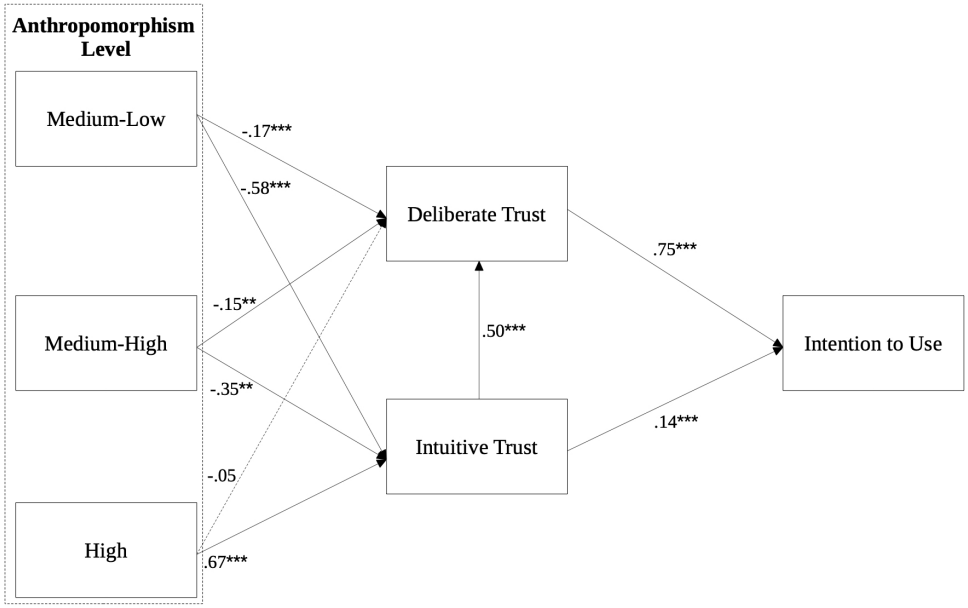

Figure 3. Results of the Mediation Model

low condition ($beta = -.05$, $p < .32$). Additionally, higher intuitive PT also lead to a higher degree of deliberate PT ($beta = -.50$, $p < .001$). A regression on intention to use showed that both deliberate PT ($beta = .75$, $p < .001$) and intuitive PT ($beta = .14$, $p < .001$) were positively and significantly related to intention to use, whereas none of the anthropomorphism conditions reached significance (all $ps > .19$). After 5000 iterations, the confidence interval for the average causal mediation effects (ACME) did not include zero for the medium-low condition ($beta = -.13$, $CI[-.21; -.05]$, $p = .002$) and the medium-high condition ($beta = -.11$, $CI[-.20; -..04]$, $p < .001$) but did include zero for the high condition ($beta = -.04$, $CI[-.12; .04]$, $p = .32$). Therefore, our results support that increasing anthropomorphism of a robot to a medium level decreases use intention through deliberate PT, whereas there is no direct effect on deliberate PT of increasing the anthropomorphism to a high level.

Intuitive PT. For intuitive PT we proceeded in three steps. In step 1, we checked the anthropomorphism level $\rightarrow$ intuitive PT $\rightarrow$ deliberate PT mediation, in step 2 the intuitive PT $\rightarrow$ deliberate PT $\rightarrow$ use intentions relationship and finally, in step 3, the anthropomorphism level $\rightarrow$ intuitive PT $\rightarrow$ intention to use mediation.

For step 1, the mediation analysis showed a significant mediation effect as well as a confidence interval excluding zero for all three anthropomorphism levels. Specifically, the ACME indicated that seeing the medium-low condition ($beta = -.29$,$CI[-.37; -.22]$, $p < .001$) and the medium-high condition ($beta = -.18$,$CI[-.25; -.11]$, $p < .001$) lead to reduced deliberate PT because of reduced intuitive PT for these robots. On the other hand, the high condition increased deliberate PT through intuitive trustworthiness ($beta = .33$,$CI[.26; .41]$, $p < .001$). For the medium-low condition ($beta = -.17$,$CI[-.28; -.07]$, $p = .002$) and the medium-high condition ($beta = -.16$,$CI[-.26; -.05]$, $p = .004$), a direct effect remained, whereas no direct effect could be detected for the high condition ($p = .31$). Therefore, especially medium anthropomorphism levels lying in the uncanny valley seem to negatively mediate PT perceptions.

2. For step 2, the mediation analysis revealed that deliberate PT mediated the positive relationship between intuitive PT and use intentions ($\beta = .38, CI[.34; .42], p < .001$), while maintaining a direct effect ($\beta = .14, CI[.10; .18], p < .001$). Thus, deliberate PT seems to have a higher effect on intention to use, while there is still an effect of intuitive PT on use intentions which is not mediated by deliberate PT.

3. For step 3, there were mediation effects for the medium-low condition ($\beta = -.08, CI[-.11; -.05], p < .001$), the medium-high condition ($\beta = -.05, CI[-.07; -.03], p < .001$), and the high condition ($\beta = .09, CI[.06; .13], p < .001$), while none of the direct effects were significant (all ps $> .190$).

5. Discussion

The following discussion of our results will be divided into two main aspects. First, we discuss the influence of anthropomorphism on intuitive and deliberate first impression PT and its consequences on further human-robot interactions. Second, we focus on our dual-processing perspective on PT for human-robot interactions and derive a resulting process model of PT evaluations and their impact on use intentions.

5.1. The Power of Anthropomorphism, or Not?

Despite social robots showing increasingly higher levels of anthropomorphic appearance in recent years, our results indicate that these levels may still be insufficient. That is, although we included highly anthropomorphic robots such as Sophia and Mark1 and participants could see them merely as images (therefore avoiding the possibility that insufficient speech production or behavior reduces anthropomorphism), the uncanny valley was still entered both for intuitive and deliberate PT, and, consequently, use intention. For PT ratings, we could observe that intuitive ratings were consistently lower than deliberate PT ratings, which might point to the reflective system consistently re-processing prior perceptions but positively toward use intentions. This assumption is also supported by our mediation model investigating the mediating effects of intuitive PT on deliberate PT, and both PT variants on use intentions. Whereas both the medium-low and medium-high conditions reduced use intentions through both deliberate and intuitive PT compared to the low anthropomorphism condition, a different picture emerged for the high anthropomorphism condition. Specifically, when comparing the high anthropomorphism condition with the low anthropomorphism condition, we could find no support that deliberate PT was able to explain relevant variance in addition to the intuitive PT. Consequently, our results indicate that deliberate PT entered the ”same” uncanny valley as intuitive PT which might be due to intuitive PT itself.

Further following this argument, when considering only the low and high anthropomorphism conditions, an increase in PT could be found for the high anthropomorphism condition. Since this increase is significant for intuitive PT, and further the mediation effect of deliberate PT was not significant for the high condition, the intuitive PT evaluation already seemed to dictate how trustworthiness is to be perceived. Thus, it is assumed that not much re-processing in the deliberate system was necessary. That is, the first impression and thus, the first intuitive evaluation seem to be consistent with the further evaluation for both the obvious, low anthropomorphic robot, and the real human, while this seems not necessarily to be the case for robots imitating humans. A result which challenges our prior assumptions that the intuitive PT evaluation might not be able to directly detect the medium-high robot as robot. An explanation as to why this effect can be observed can be given by neuroscientific studies which show that the neural processing of robots having human or human-like faces requires more cognitive effort than processing real humans or obvious robot faces [15]. This phenomenon further supports our application of a dual-processing perspective as it reflects (unconscious) decision conflicts and errors within the perception and evaluation process of social robots which might occur primarily in deliberate PT.

When further applying this finding to the design of social robots, it may need to be questioned whether we should design robots like humans. This thought has already been addressed by prior literature, suggesting that the closer the design of a robot gets to a human, the more we constrain the robot’s capabilities to those of humans [28]. Furthermore, designing robots like humans also elicits increased social processing and categorizing of these robots similar to human agents, which might lead to severe disturbances in human-robot relationships and further result in increased decision conflicts within the human brain. Consequently, it can be argued that designing robots more machine-like, but still in a way that they are perceived as trustworthy seems to be more reasonable than the aim for high anthropomorphism and creating robots indistinguishable from humans. This thought is further supported by literature stating that robots should be designed according to the tasks they will fulfill and the context in which they will interact with humans [20, 27]. Consequently, it might be reasonable to establish design guidelines focusing on social robots within a specific role each - for instance, social robots applied in elderly care might need to meet different requirements than robots which act as language teachers [47].

5.2. The Dual-Processing Perspective on Perceived Trustworthiness

The application of a dual-processing perspective of first impression PT has given us several insights into the formation and influence of PT evaluations on further use intentions. During our mediation analysis, we have shown that intuitive PT acts as mediator between the perceived anthropomorphism level of the robot and

the formed deliberate PT. Further, deliberate PT has significant mediating effects between first intuitive PT evaluations and use intentions. Nevertheless, it has been shown that a part of use intentions is also explainable with intuitive PT only. As a result, both evaluation systems might be crucial to consider when conclusions about robot acceptance and use intentions are to be drawn. These use intentions further significantly impact whether humans will actually be interacting with the robot (again) or not. The described procedure represents the main finding of our paper regarding a dual-process approach of trustworthiness perceptions in human-robot interactions and is depicted in the following Figure 4.

Therefore, applying the dual process theory of social reasoning to PT evaluations of robots allows to receive deeper insights into PT formation; and how these mediate the process from the perceived anthropomorphic level of the robot to the use or interaction intentions. For instance, we detected that the impact of anthropomorphism in the real human condition on intuitive PT is especially high and thus, we assume that there is not much re-processing in the deliberate system necessary. In line with this, we showed that deliberate PT seems to more severely impact use intention than intuitive PT. While only the first impression and appearance of robots was tested in this work, this aspect might increase in meaningfulness if actual interaction is investigated. For instance, in case the actual behavior of the robot would be worse than expected, evaluations in the deliberate system might significantly decrease which will further decrease the intention to use and interact with the robot.

6. Conclusion

6.1. Summary and Main Findings

In this research work, we have taken a dual-processing theory of PT and investigated its uncanny valley and mediating effects in first impressions of social robots. Since some robots are already designed close to humans, we have focused mostly on robots with high anthropomorphism which are already operating in practice. To gain first insights into how anthropomorphism influences PT ratings, we have focused on the robots' appearance only as a first crucial perception humans get of social robots before interacting with them. Our results show that designing robots in the image of humans does not (or not yet) seem reasonable and that their design should be much more focused and concerned with the tasks the robot will fulfill in society. This might further require specific design guidelines or requirements to be satisfied which should be matched to the role and tasks of the robot.

That being said, the application of a dual-processing perspective on PT has shown us that intuitive PT has significant mediating effects between anthropomorphism level of robots, which were significantly decreased for the human-looking robots than for the machine-like robot. Further, deliberate PT has shown to have the main mediating effect on use intentions, although a part of use intentions are also explainable by intuitive PT alone. To visualize this, we have derived an abstract process model which may help future studies in this area to consider both processes of PT in human-robot interactions.

6.2. Limitations and Future Work

As every study, this research work does not come without limitations. As its main weakness, it needs to be stated that only first impressions about the visual appearance of robots were investigated in relation to PT. PT itself is, of course, a much more complex construct which might ultimately alter in case of actual human-robot interaction. Since the aim of our study was to make a first step towards investigating the uncanny valley from a dual-processing perspective, we have derived a first model. Future research could therefore investigate this in actual human-robot interactions to further validate our results. In line with this, and as already pointed out in our discussion, it might be reasonable to conduct neuroscientific or NeuroIS studies in this field to receive neural activity as further data input. The application of neuroimaging methods could also help to overcome another limitation of this paper. That is, we distinguished intuitive and deliberate PT mainly by giving time constraints and delays, and by constructing the questions for deliberate PT more complex. While this is proposed as an appropriate method to trigger the two reasoning systems in other studies [43, 44, 30], it cannot be ensured completely that we actually triggered one of the systems at a time. Therefore, by applying neuroimaging methods, further insights into the processing in the human brain can be gained which could help to overcome this potential weakness. Moreover, investigating the neural processing and potential decision conflicts related to the uncanny valley might provide further insights into humans' unconscious perceptions of social robots.

Finally, in our study we only included four different levels of anthropomorphism and female robots (except for the low anthropomorphism condition). Future work might therefore include both male and female robots, as well as more levels of anthropomorphism which might allow to identify design aspects or criteria which provide

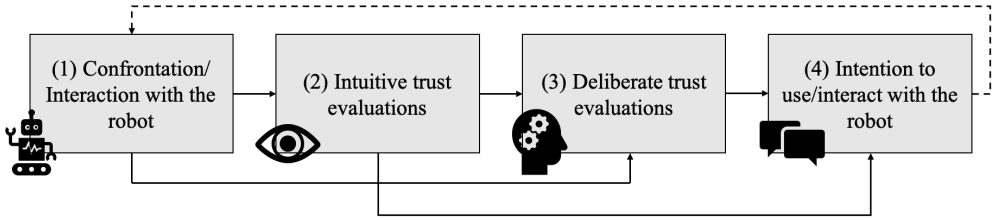

Figure 4. A Dual-Processing Model of PT

"thresholds" for when the uncanny valley is entered and when it is left. This would also provide guidance for developing design guidelines for social robots. Given this work, PT seems to be a major predictor for the uncanny valley and should therefore act as one indicator for defining these thresholds. Therewith, a dual-processing perspective on PT is recommendable to receive deeper insights into human-robot interactions.

References

[1] K. Bergmann, F. Eyssel, and S. Kopp, "A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7502 LNAI, pp. 126–138, 2012.

[2] M. Paetzel, G. Perugia, and G. Castellano, "The persistence of first impressions: The effect of repeated interactions on the perception of a social robot," in *ACM/IEEE International Conference on Human-Robot Interaction*, pp. 73–82, 2020.

[3] T. Sanders, K. E. Oleson, D. R. Billings, J. Y. Chen, and P. A. Hancock, "A model of human-robot trust: Theoretical model development," *Proceedings of the Human Factors and Ergonomics Society*, pp. 1432–1436, 2011.

[4] J. Złotowski, H. Sumioka, S. Nishio, D. F. Glas, C. Bartneck, and H. Ishiguro, "Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy," *Paladyn*, vol. 7, no. 1, pp. 55–66, 2016.

[5] D. Gefen, E. Karahanna, and D. W. Straub, "Trust and TAM in Online Shopping: An Integrated Model," *MIS Quarterly*, vol. 27, pp. 51–90, dec 2003.

[6] M. B. Mathur and D. B. Reichling, "An uncanny game of trust: Social trustworthiness of robots inferred from subtle anthropomorphic facial cues," in *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, HRI'09*, pp. 313–314, IEEE, 2008.

[7] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011.

[8] M. Lohani, C. Stokes, M. McCoy, C. A. Bailey, and S. E. Rivers, "Social interaction moderates human-robot trust-reliance relationship and improves stress coping," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 471–472, IEEE, mar 2016.

[9] M. Mori, "The uncanny valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[10] M. Strait, H. L. Urry, and P. Muentener, "Children's Responding to Humanlike Agents Reflects an Uncanny Valley," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2019-March, pp. 506–515, 2019.

[11] J. Beiboer and E. B. Sandoval, "Validating the Accuracy of Imaged-Based Research into the Uncanny Valley: An Experimental Proposal," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2019-March, pp. 608–609, 2019.

[12] K. E. Culley and P. Madhavan, "A note of caution regarding anthropomorphism in HCI agents," *Computers in Human Behavior*, vol. 29, no. 3, pp. 577–579, 2013.

[13] E. J. Lobato, T. J. Wiltshire, and S. M. Fiore, "A dual-process approach to understanding human-robot interaction," in *Proceedings of the Human Factors and Ergonomics Society*, pp. 1263–1267, 2013.

[14] F. Strack and R. Deutsch, "Reflective and impulsive determinants of social behavior," *Personality and Social Psychology Review*, vol. 8, no. 3, pp. 220–247, 2004.

[15] Saygin, A.P., T. Chaminade, and H. Ishiguro, "The Perception of Humans and Robots: Uncanny Hills in Parietal Cortex.," *Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 2716-2720). Austin, TX: Cognitive Science Society.*, 2010.

[16] Y. Wang and S. Quadflieg, "In our own image? Emotional and neural processing differences when observing human-human vs human-robot interactions," *Social Cognitive and Affective Neuroscience*, vol. 10, no. 11, pp. 1515–1524, 2014.

[17] S. Y. Komiak and I. Benbasat, "The effects of personalization and familiarity on trust and adoption of recommendation agents," *MIS Quarterly: Management Information Systems*, vol. 30, no. 4, pp. 941–960, 2006.

[18] D. S. Stoltz and O. Lizardo, "Deliberate Trust and Intuitive Faith: A Dual-Process Model of Reliance," *Journal for the Theory of Social Behaviour*, vol. 48, no. 2, pp. 230–250, 2018.

[19] I. Leite, C. Martinho, and A. Paiva, "Social Robots for Long-Term Interaction: A Survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[20] M. Lohse, F. Hegel, and B. Wrede, "Domestic applications for social robots - An online survey on the influence of appearance and capabilities," *Journal of Physical Agents*, vol. 2, no. 2, pp. 21–32, 2008.

[21] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons,

K. Snipes, A. C. Schultz, and J. Wang, "Designing robots for long-term social interaction," *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pp. 2199–2204, 2005.

[22] E. S. Kim, L. D. Berkovits, E. P. Bernier, D. Leyzberg, F. Shic, R. Paul, and B. Scassellati, "Social robots as embedded reinforcers of social behavior in children with autism," *Journal of Autism and Developmental Disorders*, vol. 43, no. 5, pp. 1038–1049, 2013.

[23] Y. Nakauchi and R. Simmons, "A social robot that stands in line," *Autonomous Robots*, vol. 12, no. 3, pp. 313–324, 2002.

[24] R. Gockley, J. Forlizzi, and R. Simmons, "Natural person-following behavior for social robots," *HRI 2007 - Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction - Robot as Team Member*, pp. 17–24, 2007.

[25] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 322–332, 2010.

[26] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler, "All robots are not created equal: The design and perception of humanoid robot heads," *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS*, pp. 321–326, 2002.

[27] K. Hayashi, M. Shiomi, T. Kanda, and N. Hagitay, "Who is appropriate? A robot, human and mascot perform three troublesome tasks," *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 348–354, 2010.

[28] B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 177–190, 2003.

[29] S. X. Komiak and I. Benbasat, "Understanding Customer Trust in Agent-Mediated Electronic Commerce, Web-Mediated Electronic Commerce, and Traditional Commerce," *Information Technology and Management*, vol. 5, no. 1/2, pp. 181–207, 2003.

[30] I. Sarmany-Schuller, "Decision Making Under Time Pressure In Regard to Preferred Cognitive Style (Analytical-Intuitive) and Study Orientation," *Studia Psychologia*, vol. 52, no. 4, pp. 285–290, 2010.

[31] S. A. Jessup, A. M. Gibson, A. Capiola, G. M. Alarcon, and M. Borders, "Investigating the Effect of Trust Manipulations on Affect over Time in Human-Human versus Human-Robot Interactions," in *HICSS*, pp. 553–563, 2020.

[32] N. K. Lankton, D. Harrison McKnight, and J. Tripp, "Technology, humanness, and trust: Rethinking trust in technology," *Journal of the Association for Information Systems*, vol. 16, no. 10, pp. 880–918, 2015.

[33] J. Elson, D. C. Derrick, and G. S. Ligon, "Trusting a Humanoid Robot: Exploring Personality and Trusting Effects in a Human-robot Partnership," in *HICSS*, (Hawaii), pp. 543 – 553, 2020.

[34] N. Pfeuffer, A. Benlian, H. Gimpel, and O. Hinz, "Anthropomorphic Information Systems," *Business and Information Systems Engineering*, vol. 61, no. 4, pp. 523–533, 2019.

[35] M. B. Mathur, D. B. Reichling, F. Lunardini, A. Geminiani, A. Antonietti, P. A. Ruijten, C. A. Levitan, G. Nave, D. Manfredi, B. Bessette-Symons, A. Szuts, and B. Aczel, "Uncanny but not confusing: Multisite study of perceptual category confusion in the Uncanny Valley," *Computers in Human Behavior*, vol. 103, no. September 2019, pp. 21–30, 2020.

[36] M. Seymour, K. Riemer, and J. Kay, "Interactive Realistic Digital Avatars - Revisiting the Uncanny Valley," *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)*, pp. 547–556, 2017.

[37] M. B. Mathur and D. B. Reichling, "Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley," *Cognition*, vol. 146, pp. 22–32, 2016.

[38] R. Riedl, M. Hubert, and P. Kenning, "Are There Neural Gender Differences in Online Trust? An fMRI Study on the Perceived Trustworthiness of eBay Offers," *MIS Quarterly*, vol. 34, no. 2, pp. 397–428, 2010.

[39] D. H. McKnight, V. Choudhury, and C. Kacmar, "Developing and validating trust measures for e-commerce: An integrative typology," *Information Systems Research*, vol. 13, no. 3, pp. 334–359, 2002.

[40] D. H. Mcknight, M. Carter, J. B. Thatcher, and P. F. Clay, "Trust in a specific technology: An investigation of its components and measures," *ACM Transactions on Management Information Systems*, vol. 2, no. 2, 2011.

[41] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User Acceptance of Somputer Technology: A Comparison of Two Theoretical Models," *Management Science*, vol. 35, no. 8, pp. 982–1003, 1989.

[42] M. Blanz, *Forschungsmethoden der Statistik für die Soziale Arbeit: Grundlagen und Anwendungen*. Stuttgart: Kohlhammer, 2015.

[43] A. Glöckner and C. Witteman, *Foundations for Tracing Intuition: Challenges and Methods*. Taylor & Francis, 2009.

[44] C. Betsch and J. J. Kunz, "Individual strategy preferences and decisional fit," *Journal of Behavioral Decision Making*, vol. 21, pp. 532–555, dec 2008.

[45] Z. Zhang, M. J. Zyphur, and K. J. Preacher, "Testing Multilevel Mediation Using Hierarchical Linear Models: Problems and Solutions," *Organizational Research Methods*, vol. 12, pp. 695–719, Oct. 2009.

[46] D. Tingley, T. Yamamoto, K. Hirose, L. Keele, and K. Imai, "mediation: R Package for Causal Mediation Analysis," *Journal of Statistical Software*, vol. 59, no. 5, 2014.

[47] T. Belpaeme, P. Vogt, R. van den Berghe, K. Bergmann, T. Göksun, M. de Haas, J. Kanero, J. Kennedy, A. C. Küntay, O. Oudgenoeg-Paz, F. Papadopoulos, T. Schodde, J. Verhagen, C. D. Wallbridge, B. Willemsen, J. de Wit, V. Geçkin, L. Hoffmann, S. Kopp, E. Krahmer, E. Mamus, J. M. Montanier, C. Oranç, and A. K. Pandey, "Guidelines for Designing Social Robots as Second Language Tutors," *International Journal of Social Robotics*, vol. 10, no. 3, pp. 325–341, 2018.