# The Attribution of Emotional State - How Embodiment Features and Social Traits Affect the Perception of an Artificial Agent

Maike Paetzel[1], Giovanna Varni[2], Isabelle Hupont[3],
Mohamed Chetouani[3], Christopher Peters[4] and Ginevra Castellano[1]

***Abstract*— Understanding emotional states is a challenging task which frequently leads to misinterpretation even in human observers. While the perception of emotions has been studied extensively in human psychology, little is known about what factors influence the human perception of emotions in robots and virtual characters. In this paper, we build on the Brunswik lens model to investigate the influence of (a) the agent's embodiment using a 2D virtual character, a 3D blended embodiment, a recording of the 3D platform and a recording of a human, as well as (b) the level of human-likeness on people's ability to interpret emotional facial expressions in an agent. In addition, we measure social traits of the human observers and analyze how they correlate to the success in recognizing emotional expressions. We find that interpersonal differences play a minor role in the perception of emotional states. However, both embodiment and human-likeness as well as related perceptual dimensions such as perceived social presence and uncanniness have an effect on the attribution of emotional states.**

## I. Introduction

Artificial agents have a great potential to become social companions in our everyday life. While platforms like Amazon Alexa and Google Home can already handle basic task-related interactions, they are still lacking the social aspect in conversations. Becoming social does not just depend on the agent's ability to engage in a natural verbal conversation; humans also rely on "expressive moods and emotions as an integral part of social interaction" [1]. Showing emotions not only conveys the agent's internal state, it is also crucial to communicate empathy [2] and to build rapport [3].

Humans are not able to directly examine the internal state of another (living) being in their environment - instead, they have to rely on observable cues to build their own model of the other's state. Egon Brunswik [4] came up with a model known as the *Brunswik lens* which describes the processes involved in perceiving an internal state of another entity (cf. Fig. 1). This model distinguishes between *distal* and *proximal* cues. Distal cues are any form of observable behavior that can be perceived by others, while proximal cues activate the attribution process, i.e. the perceptual judgment that accounts for the emotions and personality traits an observer attributes to a person being observed [5]. For example, during a face-to-face interaction, the energy of acoustic waves and facial muscle movements are the distal cues, while loudness and the perceived facial expression are the respective proximal cues. Similar to [6], we argue that the Brunswik lens is a simple but powerful model of human cognition that can also be applied to human-machine perception.

While human psychology has already addressed the mechanisms of attributing emotional states to another human [7], little is known about what influences the attribution process that leads to a perceived (emotional) state of an artificial agent. In this paper, we focus on investigating how people attribute the six basic emotions to an artificial agent [8]. We manipulate the distal cues of the agent by using different *embodiments* and modifying the *virtual facial muscle activations* (Action Units [9]) and *level of human-likeness in the texture* of each embodiment, because the overall appearance of human-likeness and the embodiment has already been shown to influence the perception of an agents' (non-)human emotional capabilities [10][11][12]. In this work, we use the 3D blended embodiment *Furhat* [13] (cf. Fig. 1, left) which can convey human-like expressions in a smooth and noiseless way. We compare it with a virtual face, a video recording of the 3D embodiment and a video recording of a human.

By analyzing the correlation between the different perceived traits of an artificial agent, our research contributes to the understanding of how humans perceive agents in social interactions. While most related work is focused on artificial agents as observers or the externalization process of an agent's internal state, we investigate the *underlying cues that lead a human observer to attribute certain states to an artificial agent*. Therefore, we are not only measuring the perceived emotional state of the agent; we are also interested in understanding how the manipulated distal cues as well as social traits of the observer itself influence the perceived *human-likeness*, *social presence* and *level of uncanniness*.

## II. Related Work

The Brunswik lens is a simple but well-established model from social psychology to investigate human-human interactions and emotions. It gained attention in recent years again when it was adapted to human-agent interactions to study how humans perceive and behave around agents. Vinciarelli's work [6] is focused on social signal synthesis and the automated processing of social signals by machines, omitting the topic of human cognitive processes when attributing behavior to artificial agents. However, a deeper investigation

[1]Uppsala University, Department of Information Technology, Sweden {maike.paetzel, ginevra.castellano}@it.uu.se
[2]LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France giovanna.varni@telecom-paristech.fr
[3] Sorbonne Université, Institut des Systèmes Intelligents et de Robotique, CNRS UMR7222 Paris, France {hupont,chetouani}@isir.upmc.fr
[4]KTH Royal Institute of Technology, Department of Computational Science Technology, Sweden chpeters@kth.se

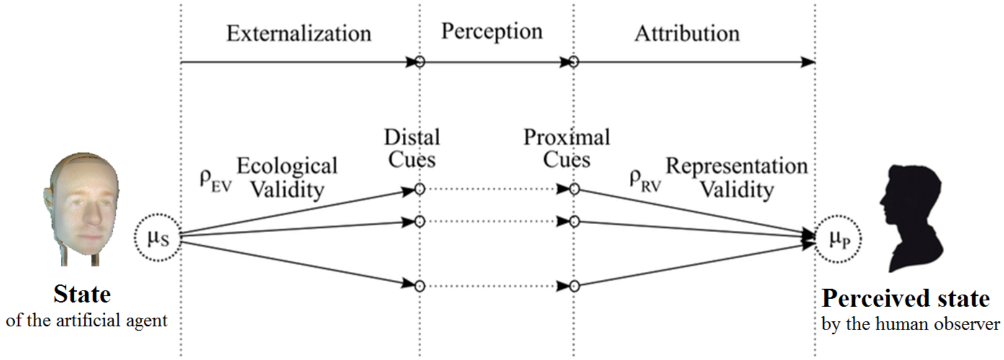


Fig. 1: Adaptation of the Brunswik lens model [6] in which a human observer perceives the internal state of a robot.

of those processes has the potential to improve human-agent interaction in the future. For example, [14] found that cues humans use for their own behavioral synthesis only overlap to some extent with those used for attribution of behavior by others, which can lead to errors in the attribution process. We believe that a better understanding of the cues humans use to interpret the behavior of artificial agents will lead to fewer errors in human-agent interaction in the future.

In this paper, we aim to further understand the attribution side of the Brunswik lens by investigating influencing factors on the perception of emotional state. [15] introduced the idea that accurate encoding models of emotions (those that "faithfully represent how people encode emotion") are not necessarily accurate decoding models (which "maximize the likelihood that an observer correctly interprets the emotional signal"). Apart from different approaches to synthesizing expressions, the embodiment has been shown to influence the ability of humans to recognize emotions. There are opposing findings regarding whether the emotions of virtual agents and robots are more difficult to recognize than human emotions, potentially because of the variety of robot platforms used. [11] found an overall indication towards robot platforms being more difficult to interpret, especially for negative expressions. This is in line with previous studies which have demonstrated that negative emotions are the hardest to decode even in human-human interaction [16]. Blended embodiments like Furhat [13] are comparably new platforms and little is known about how findings in fully physical or virtual embodiments may translate to them. By comparing a blended embodiment with a fully virtual version of the same face, our work contributes to the general understanding of the perception of such embodiment types.

While the differences in expressivity of virtual and physical embodiments might explain the varying ability of humans to interpret facial expressions, it could also be related to the agents' social presence [17]. It is important to note that social presence does not equal physical embodiment. Even though physical embodiments can increase social presence [18] it is not necessarily implied and little is known about the relation between them [19].

In addition to the embodiment, the level of human-likeness has shown to generally influence the perception of an artificial agent: People have more empathy towards anthropomorphic robots [20], but an appearance that is too anthropomorphic might increase feelings of uncanniness [21]. [10] found that the more human-like an agent looks, the higher its anticipated emotional capabilities are. However, the influence of human-likeness on the ability to attribute emotional state to an artificial agent has not been investigated so far.

The perception of emotional state depends not only on the entity being observed but also on the observer. The link between the observer's personality traits and perception of basic emotions in another human has already been addressed in neuroscience. [22], for example, showed a significant positive correlation between the observer's level of extroversion and perceived happiness in others. [23] looked at the extroversion level of an observer and its influence on the perception of emotional cues in a human, two robot platforms and a virtual human. They found that extroversion did not affect the recognition rate. However, people with low extroversion scores were better than those with high extroversion scores at recognizing emotions acted by less visually complex characters. An open question is how other personality or social traits of an observer influence the perception of emotional state in an artificial agent.

## III. Research Questions

With our work we aim to gain further insights in the distal and proximal cues that influence the perception of an artificial agent's emotional state. Emotional expressions are an essential means for artificial agents to convey internal states and believes, and therefore a main building block in shaping social conversations [24]. To understand the influencing factors when attributing emotional state, we manipulate three types of cues in the agent: the agent's facial expressions, embodiment and level of human-likeness. We state the following research question (RQ):

**(1a)** Do the agent's embodiment and level of human-likeness influence the perception of the agent's emotional state?

In addition, we investigate the influence of the observer's social traits, specifically the observer's personality and level of empathy, on the perception of the agent:

**(1b)** Do the observer's social traits influence the perception of the agent's emotional state?

By controlling the level of human-likeness and embodiment, it is likely that we are influencing factors beyond the attribution of an emotional state to the agent. As discussed in Sec. II, the literature indicates a strong correlation between the level of human-likeness and agent's likability (or perceived level of uncanniness) as well as social presence. To further understand the influencing factors related to the attribution of emotional state, we need to understand the correlation between the emotional state and other perceived traits in the artificial agent. We state the following research questions:

**(2a)** Does the agent's embodiment, level of human-likeness and the observer's social traits influence the perception of the agent's uncanniness and social presence?

**(2b)** Are the perceived level of uncanniness and social presence correlated with the ability to attribute emotional state to an artificial agent?

## IV. Methodology

We designed an experiment with the following independent variables:

- *Type of embodiment* with the four levels: Furhat 3D blended embodiment [13], video recording of Furhat robot, 2D virtual character, video recording of a human (within-subjects, cf. Fig. 2 top row)
- *Level of human-likeness* with the three levels: human-like, morph and character-like (between-subjects, cf. Fig. 2 bottom row)

The agent's emotional state was manipulated using the facial expressions of the six basic emotions: happiness, sadness, surprise, fear, anger and disgust [8]. Emotional facial expressions were presented in short sequences including onset, apex and offset, without vocalizations or head movements (cf. Fig. 3). The perceived emotional state and the perceived human-likeness, social presence and uncanniness of the agent are measured as dependent variables in the experiment. In addition, we collect information about participants' personality traits and level of empathy.

### A. Participants

46 students (11 female; age *M*=26.20, *SD*=4.33; 20 Swedish) were recruited from a course at Uppsala University to participate in the experiment. All participants have at least a high school degree and have been or are currently enrolled in a computer science program. Participants had mostly advanced or fluent English language skills. They were evenly distributed within the three conditions (16 in the human-like condition and 15 each in the character-like and morph condition).

### B. Apparatus and Stimuli

A male virtual face was presented to the participants in three different embodiments with approximately the same size: A *2D representation on a screen*, a *3D back-projected version on a Furhat* robot [13] and a *video-recording of the Furhat* robot (cf. Fig. 2 top row). A *video recording of a real human* was used as the fourth embodiment. The virtual face and the two video recorded embodiments were presented on a screen in portrait orientation. The Furhat robot is a 3D blended embodiment platform equipped with a rigid mask of a male face on which the texture is projected from within.

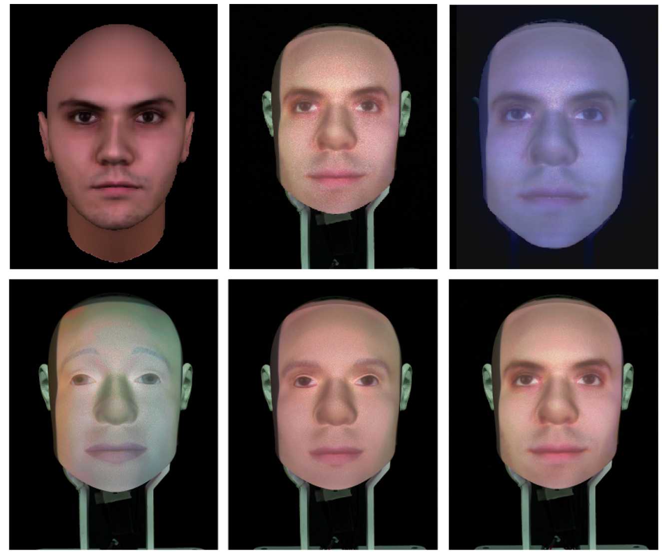

Fig. 2: Top left to right - Three artificial embodiments: 2D virtual agent on screen, 3D bended embodied robot Furhat, video recording of 3D embodiment.
Bottom left to right - Levels of human-likeness: Character-like, morph and human-like.

The exact same stimuli were synthesized for the virtual face and the Furhat robot. Two online pre-studies were carried out to select the best facial textures and facial animations to use in the main study, respectively, as described below.

*1) Synthesis of facial textures:* The human-likeness of the character was controlled using three different male virtual facial textures which were selected in a pre-study on Amazon Mechanical Turk. In total, we created 28 candidate textures: *human-like* faces were created in FaceGen Modeller [25] from pictures of real human faces and the *character-like* faces were standard synthetic Furhat faces. The *morphs* between the human- and character-like textures were created by blending between skin textures in the Paint.NET digital photo editing package. The face was then projected onto the Furhat robot and pictures were taken from the robot in each configuration.

20 crowd workers were recruited to participate in this first pre-study. For each texture, they were asked to judge how human-like the character is (2 questions based on the Godspeed questionnaire [26]) as well as how likable and threatening it appears (3 questions each [27]). Three final textures were selected from the set of stimuli: A character-like version (cf. Fig. 2, bottom row, left), a human-like version (cf. Fig. 2, bottom row, right) and a morph between the two (cf. Fig. 2, bottom row, center). They were chosen on the basis of crowd workers' ratings so that the human-like texture was rated more human-like and likable and less threatening than the other two versions, and the morph was rated in between the human-like and character-like version for all three dependent variables anthropomorphism, likability and perceived threat.

Fig. 3: Synthesis of the six basic facial emotions: happiness, sadness, surprise, fear, anger and disgust (from left to right).

*2) Synthesis of facial animations:* In our main study, all embodiments expressed the six basic emotions [8] to the participants. We selected a video recording of a male performing posed facial expressions from the MUG database [28]. He was following the Facial Action Coding System (FACS) [9] when performing the six basic emotions. Our synthesis of facial animations for the artificial characters replicate the dynamics of the human video recording as closely as possible. However, in the IrisTK animation system [29] some facial Action Units are combined and cannot be controlled separately, which made it necessary to have small differences between the human expressions and the artificial synthesis. While an expert in the FACS coding system ensured that the synthesis was still following the FACS guidelines, we could not be certain how the differences would change the perception of the facial expressions. Specifically, we were aiming to ensure that all facial expressions have the same recognition rate, intensity, naturalness and strangeness among the 2D and 3D embodiment and the human video recording.

To understand perceptual differences between the embodiments, we designed a second pre-study in which facial expressions of the six basic emotions were performed by the Furhat 3D blended embodiment, the 2D virtual character and a human. Short video clips of each expression and embodiment were recorded. Every crowd worker rated all facial expressions (within-subject) for one of the embodiments (between-subject). For each expression, participants were asked to select which expression was displayed and how intense, natural and strange each emotion was perceived.

60 crowd workers participated in this pre-study and were evenly distributed into the three embodiment conditions. For most expressions, participants were able to ascribe the intended emotional state to the shown stimulus. Surprise and happiness were correctly identified by almost all participants in all embodiments, while the negative emotions were more difficult to identify. This is, however, not unexpected and in line with previous studies showing that facial expressions of negative emotions are generally more difficult to decode [16]. Errors in the attribution process were mostly coherent between the embodiments, with a small increase in the recognition rate for sadness and anger in the two artificial characters. By achieving human-level recognition rates and better, our synthesis can be considered successful.

We performed a 3x6 ANOVA analysis with embodiment as between-subjects factor and emotion as within-subjects factor and Bonferroni adjusted alpha levels of 0.004 (0.05/12). The level of intensity was comparable in all expressions except for anger where the 2D and 3D synthesis was rated more intense than the human, $F(2,48) = 51.5, p < .001$. We could not find a significant difference in the naturalness and strangeness of all emotional expressions except for anger, which was more strange in the artificial synthesis, $F(2,48) = 13.4$, $p < .001$. Since there was no systematic difference between the artificial synthesis and the human stimulus, the synthesis was considered sufficient for the on-site experiment.

### C. Questionnaires

The observer's individual traits were measured with questionnaire (Q1) which was handed out on a laptop before the start of the experiment. (Q1) includes a general demographics questionnaire (10 items), the short version of the big 5 personality traits (10 items) [30] and the Interpersonal Reactivity Index (IRI, 21 questions, excluded personal distress, Cronbach's $\alpha$ between .70 and .78 according to [31]).

Questionnaire (Q2) was shown after each facial expression to assess the perceived emotional state of the agent. It is composed of the single-choice question "Which of these facial expressions was just displayed?" with the six basic emotions and "neutral" as response options and the question how certain they were in their selection with the three options: "Uncertain", "Neither nor", "Certain".

Questionnaire (Q3) was shown with the same order of questions and items after every embodiment to measure participant's general assessment of the character on three different levels, all assessed on 5-point Likert scales:

- *Level of Anthropomorphism* (5 items), sub-scale from the Godspeed questionnaire [26], Cronbach's $\alpha = .91$ according to [32].
- *Social presence* (8 items), excerpt from the social presence questionnaire by [33], Cronbach's $\alpha \geq 0.82$ for all sub-scales the questions were taken from.
- *Uncanniness* (10 items), excerpt from [27], Cronbach's $\alpha \geq 0.82$ for the two relevant sub-scales likability and threat.

### D. Experimental Setup & Procedure

The experimental sessions took place in a laboratory room at Uppsala University, Sweden. After arriving at the session, participants were informed about the experimental procedure, signed a consent form and answered questionnaire Q1 on a laptop. They were then seated at a distance of about 100 cm from an artificial character and instructed to watch the character displaying facial expressions. The character was placed at approximately 100 cm from the ground, which was roughly on eye level for the majority of participants.

Each emotional stimulus started and ended with a beep tone. After the second beep tone, participants were told to select which facial expression was displayed in questionnaire Q2 on an iPad placed at the table in front of them. Once they finished Q2, the character automatically generated the beep tone followed by the next facial expression after a pause of about 2 seconds. The procedure was repeated until each facial expression was displayed twice. The order of the expressions was randomized with the constraint that no expression was displayed twice consecutively. After all 12 expressions (2 x 6 emotions) were finished for one embodiment, participants were guided to fill out Q3 on the iPad. In the meantime, the embodiment type was changed by the researcher if necessary. Once participants responded to Q3, they followed the same procedure for the next embodiment without a break.

## V. Results

People attributed the intended emotional state for the majority of trials, with an overall match of 77% between the synthesized state $\mu_S$ and perceived emotional state $\mu_P$. With 50% difference between $\mu_S$ and $\mu_P$, fear was the expression that was most often confused with another emotional state, mostly sadness for the artificial synthesis and surprise for the human video. The difference between $\mu_S$ and $\mu_P$ was comparably high over all four embodiments, including the human video.

The second lowest match between $\mu_S$ and $\mu_P$ was observed in sadness, which was recognized correctly in only 51.9% of all trials. This was specifically due to the failure to attribute the intended emotional state in all artificial embodiments. A similar phenomenon is seen in disgust, where the intended emotional state in the human is attributed in 95% of all cases, while the recognition rate among the agent embodiments is as low as 75% for 2D and 83% for 3D. Interestingly, for anger this trend is reversed: Participant's success rate in identifying this emotion was 87% for the 3D mixed embodiment and the 2D version of the artificial character, but only in 76% of the trials this emotion was correctly attributed to the recording of the human. The differences between $\mu_S$ and $\mu_P$ in anger can be explained by the findings in our online pre-study, which revealed that anger was displayed more intensely in the artificial synthesis. The best overall recognition rates are achieved in happiness (94.8%) and surprise (98%), which is again in line with the findings in our pre-study and the literature. A confusion matrix with the displayed emotion versus the selected emotions summarized over all embodiment and human-likeness conditions is shown in Table I.

The human-likeness conditions influenced people's perception as intended: A 3x4 ANOVA with level of human-likeness (human-like, morph, character-like) as between-subjects factor and embodiment (3D blended embodiment, video recording of 3D embodiment, 2D virtual character, video recording of a human) as within-subjects factor and Bonferroni adjusted alpha levels of .004 per test (.05/12) showed that the human-like condition was perceived as significantly more human-like (*M*= 3.40, *SD*= 0.11) compared to the morph (*M*= 2.94, *SD*= 0.14), $p = .001$. It was also rated as more human-like compared to the character-like condition (*M*= 3.05, *SD*= 0.12), $p = .023$, although this trend is not significant with Bonferroni adjusted alpha levels. Similarly, the video-recording of the human (*M*= 4.14, *SD*= 0.11) is perceived as significantly more human-like than the 3D mixed embodiment (*M*= 3.10, *SD*= 0.11), $p < .001$, the 2D character (*M*= 2.60, *SD*= 0.13), $p < .001$, and the video-recording of the 3D character (*M*= 2.70, *SD*= 0.11), $p < .001$.

TABLE I: Confusion matrix of the displayed vs. selected emotional state combined for all embodiment and human-likeness conditions.

| Emotional state $\mu_S$ | Perceived emotional state $\mu_P$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | ANG | DIS | FEA | HAP | SAD | SUR | NEU |
| ANG | **314** | 34 | 3 | 0 | 11 | 0 | 6 |
| DIS | 47 | **310** | 2 | 0 | 0 | 5 | 4 |
| FEA | 1 | 17 | **175** | 0 | 89 | 84 | 1 |
| HAP | 0 | 1 | 0 | **349** | 1 | 1 | 16 |
| SAD | 44 | 30 | 6 | 0 | **191** | 5 | 92 |
| SUR | 0 | 0 | 3 | 0 | 2 | **361** | 2 |

### A. RQ1: Perception of Emotional State

RQ1 is dedicated to investigate the influence of (a) the controlled distal cues human-likeness and embodiment and (b) the observer's social traits on the perception of emotional state. A 3x4 ANOVA with level of human-likeness as between-subjects factor and embodiment as within-subjects factor and Bonferroni adjusted alpha levels of .004 per test revealed main effects of embodiment, $F(3, 172) = 7.353, p < .001$, and human-likeness, $F(2, 172) = 10.891, p < .001$, on the ability to correctly attribute the intended emotional state to the stimulus.

A Tukey's Post-Hoc test showed the percentage of correctly attributed emotions to be significantly higher in the character-like version (*M*= 0.83, *SD*= 0.01) compared to both the human-like condition (*M*= 0.73, *SD*= 0.02), $p < .001$, and the morph (*M*= 0.75, *SD*= 0.02), $p = .001$. People had the highest success rate in attributing the intended facial expressions to the video-recording of the human (*M*= 0.84, *SD*= 0.02) compared to the 3D (*M*= 0.74, *SD*= 0.02), $p < .001$, and 2D version (*M*= 0.74, *SD*= 0.02), $p < .001$, and the video-recording of the 3D character (*M*= 0.76, *SD*= 0.02), $p = .01$. There is no significant difference in the ability to attribute facial expressions to the three versions of the artificial character.

A Spearman's rank correlation analysis did not show a significant correlation between participant's social traits (personality and level of empathy) and their ability to attribute emotional state to the stimulus. Only participant's level of extroversion is negatively correlated with people's success in attributing the intended emotion ($\rho = -0.219, p = .003$).

TABLE II: Excerpt of the correlation between people's perception of a social agent and their personality traits. Significance in correlations is indicated as $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

| Personality Trait | Perception | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Threat | Likability | Anthropomorphism | Emotional Interdep. | Behavioral Interdep. | Co-Presence | Attention Allocation | Affective Understanding |
| Agreeableness | | 0.184 (*) | | | | 0.217 (**) | 0.22 (**) | 0.247 (***) |
| Conscientiousness | | | | -0.241 (**) | -0.391 (***) | | | |
| Openness to Experiences | | | | | | -0.18 (*) | -0.162 (*) | -0.161 (*) |
| Fantasy | | | | 0.171 (*) | | 0.169 (*) | | |
| Empathic Concern | | | -0.145 (*) | 0.159 (*) | | -0.168 (*) | -0.164 (*) | |
| Perspective Taking | | | | 0.244 (***) | 0.172 (*) | | | |

## B. RQ2a: Perception of Uncanniness and Social Presence

RQ2 consists of two parts: First, we are interested in understanding how the controlled distal cues and the observer's internal state influence the perceived uncanniness and social presence of the agent in our particular study setup. In the following section we then use those findings and investigate if there are correlations between the attribution of emotional state and the perceived uncanniness and social presence of the agent.

*1) Perceived Uncanniness:* The perceived uncanniness of the agent was measured using questions about the perceived threat and likability of the agent. A 3x4 ANOVA with level of human-likeness as between-subjects factor and embodiment as within-subjects factor and Bonferroni adjusted alpha levels of .004 per test revealed a main effect of embodiment, $F(3, 172) = 7.486, p < .001$, and no effect of human-likeness, $F(2, 172) = 0.796, p = .453$, on the perceived threat of the agent. A Tukey's Post-Hoc test showed that the video recordings were generally perceived as less threatening than the live agents: The 2D agent (*M*= 2.12, *SD*= 0.11) was perceived significantly more threatening than the video recording of the 3D embodiment (*M*= 1.77, *SD*= 0.08), $p = .042$, and the video recording of the human (*M*= 1.51, *SD*= 0.07), $p < .001$. The 3D blended embodiment (*M*= 1.88, *SD*= 0.10) was perceived as significantly more threatening than the recording of the human as well, $p = .026$.

The same 3x4 ANOVA analysis also revealed main effects of embodiment, $F(3, 172) = 9.401, p < .001$, on the perceived likability of the agent. The embodiment has an additional influence on the perceived human-likeness, $F(2, 172) = 4.707, p = .01$. However, this trend is not significant with respect to the corrected alpha level. A Tukey's Post-Hoc test found the human-like version (*M*= 2.83, *SD*= 0.09) to be liked significantly better than both the character-like version (*M*= 2.49, *SD*= 0.10), $p = .032$, and the morph (*M*= 2.47, *SD*= 0.11), $p = .02$. The video-recording of the human (*M*= 3.03, *SD*= 0.11) is liked significantly better than the 2D character (*M*= 2.27, *SD*= 0.11), $p < .001$, and the video-recording of the 3D embodiment (*M*= 2.43,*SD*= 0.11), $p < .001$. It is not significantly more liked than the 3D mixed embodiment (*M*= 2.68, *SD*= 0.10), $p = .1$. However, the 3D mixed embodiment has a significantly higher rating than the 2D embodiment, $p = .041$.

*2) Perceived Social Presence:* Using a 3x4 ANOVA with level of human-likeness as between-subjects factor and embodiment as within-subjects factor and Bonferroni adjusted alpha levels of .004 per test we found a main effect of human-likeness on the perceived co-presence of the agent, $F(2, 172) = 9.407, p < .001$. There is an additional influence on the perceived emotional interdependence, $F(2, 172) = 4.069, p = .019$. However, this trend is not significant with respect to the corrected alpha level. A Tukey's Post-Hoc test found the emotional interdependence with the character-like version (*M*= 2.28, *SD*= 0.13) to be rated significantly higher than the human-like version (*M*= 1.83, *SD*= 0.11) and higher, though not significantly, than the morph (*M*= 1.90, *SD*= 0.13). The co-presence of the morph (*M*= 3.51, *SD*= 0.12) was rated significantly lower than both the human-like (*M*= 3.98, *SD*= 0.08) and character-like version (*M*= 4.08, *SD*= 0.09). The embodiment has a main effect on the affective understanding of the agent, $F(3, 172) = 8.157, p < .001$. Influences on the emotional interdependence, $F(3, 172) = 4.116, p = .008$, and co-presence, $F(3, 172) = 4.224, p = .006$, are not significant with respect to the corrected alpha levels. The video-recording of the human has generally the highest social presence of the embodiments, but the effect is only significant towards all artificial embodiments in the scale of affective understanding. For emotional interdependence and co-presence, the effect is only significant towards the video recording of the 3D embodiment.

A Spearman's rank correlation analysis did not show a significant correlation between the perceived social presence of the character and the perceived threat. However, the likability has a significant positive correlation with the perceived social presence of the character (Emotional interdep.: $\rho = 0.318, p < .001$; Behavioral interdep.: $\rho = 0.301$, $p < .001$; Co-Presence: $\rho = 0.224, p = .002$; Attention Allocation: $\rho = 0.16, p = .002$; Affective Understanding: $\rho = 0.365, p < .001$). Similarly, the perceived anthropomorphism is significantly correlated with the perceived social presence (Emotional interdep.: $\rho = 0.263, p < .001$; Behavioral interdep.: $\rho = 0.329, p < .001$; Co-Presence: $\rho = 0.255, p < .001$; Attention Allocation: $\rho = 0.172$, $p = .03$; Affective Understanding: $\rho = 0.426, p < .001$). In summary, the more likable and human-like a character was perceived, the more socially present it was rated.

A Spearman's rank correlation test was performed between peoples' perception of the agent and their own personality. Results are summarized in Table II. They show that the perception of uncanniness and anthropomorphism is mostly independent from participant's social traits. However, the perceived social presence is related to their agreeableness, conscientiousness, openness to experience and level of empathy.

### C. RQ2b: Correlation between Perceived Traits

To assess the correlation between people's perception of the robot based on the observed distal cues and their attribution of emotional state, a Spearman's rank correlation test was performed. It shows a significant negative correlation between the perceived threat and the success in attributing the intended emotional state ($\rho = -0.207, p = .005$), as well as a positive correlation between the task success and the perceived likability of the agent ($\rho = 0.182, p = .013$). This suggests that the perception of uncanny indeed decreases people's ability to recognize emotions in an artificial agent. Similarly, the perceived anthropomorphism correlates with the ability to recognize emotional state ($\rho = 0.183, p = .013$). In addition, a perceived higher social presence of the agent correlates with people's ability to recognize the agent's emotional state. Specifically for the co-presence ($\rho = 0.222, p = .002$), the attention allocation ($\rho = 0.185, p = .012$) and the affective understanding ($\rho = 0.313, p < .001$) the effect is significant.

## VI. Discussion

The distal cues we manipulated influenced people's ability to attribute emotional state to the stimuli (RQ1a):

*1) Facial Expressions:* Facial expressions of negative emotions like sadness or fear were more difficult to interpret compared to the positive emotions happiness and surprise. This finding is in line with our online pre-study and the related literature [16].

*2) Embodiment:* In contrast to our online pre-study, the main experiment showed that people are significantly better in attributing the intended emotional state to a recording of a real human compared to an artificial agent, even though the synthesis of facial expressions was based on the exact same human recording. However, our results also suggest that specific syntheses have the potential to surpass human recognition rates, which would make artificial agents easier to interpret than a human interlocutor while keeping the level of expressiveness. This further supports the theory that accurate encoding models are not necessarily accurate decoding models and a higher success in attributing emotional state could likely be achieved by using a better optimized decoding model [15].

In this work, we used a 3D blended embodiment and could show that the *attribution of emotional state to this blended embodiment is as successful as the attribution to a state-of-the-art virtual face*. This finding is relevant because it gives initial task-specific insights into how blended embodiments compare to virtual embodiments of an artificial agent. Future work should build upon this and investigate how these findings relate to a full 3D embodiment.

*3) Human-Likeness:* It might sound intuitive that the closer an artificial face resembles the face of a real human, the closer it is to what we are used to and thus the easier it is to attribute emotional state. However, our findings suggest the opposite: Among our three control conditions, people are most successful in attributing emotional state to the character-like version. A possible explanation is the higher level of abstraction in the character-like face which potentially helps examining the emotional facial expressions. Nevertheless, the correlation test indicated a significant positive correlation between the perceived anthropomorphism and the ability to attribute emotional state to the agent. This correlation is, however, highly influenced by the video recording of the real human and close to zero if only the three artificial agents are taken into account. More insights can be gained by including characters which are rated more towards the extremes of the anthropomorphism scale in future work.

While RQ1a aimed to investigate whether a general influence of distal cues on the attribution of emotional state exists, RQ2b was focused on understanding correlations between different attributed states. Our results indicate that different perceived traits of an agent indeed confound the ability to recognize emotional expressions. More specifically, we find *the more socially present a character is perceived and the lower its perceived uncanniness is, the more successful people are in attributing the intended emotional state to it.*

Interpersonal differences between observers that we investigate with RQ1b and RQ2a play only a minor role in the perception of emotional expressions in an agent. Unlike [23], we did find a negative influence of the observer's extroversion level on the ability to correctly identify emotional facial expressions. This might be due to the fact that the majority of our expressions were negative while extroverts have been shown to have higher amygdala responses to positive expressions which are more common for their own personality [22]. The *observer's personal traits also influence the perceived social presence of the agent*: People with a higher level of conscientiousness are more consistent and self-disciplined in their behavior and might therefore be less interdependent from and affected by the agent's behavioral and emotional state. People with a higher openness to experiences are more objective and (self-)reflected, which could explain why they judge the social presence of the character to be lower in general. Our findings further indicate that *interpersonal differences do not influence the perceived human-likeness and level of uncanniness of an agent*. This finding is very relevant for related research on the uncanny valley effect, because it suggests that the observer's personality traits and level of empathy do not confound obtained results.

Our work is also relevant for uncanny valley related research as it suggests that *video recordings of an agent are perceived as less threatening compared to a live agent interaction*, even if that agent only has a virtual embodiment. Thus, related work which often makes use of video recordings instead of live interactions might not be easily transferable. The relation between the perceived social presence and perceived uncanniness of the agent is complex and requires further investigations. Our results indicate that the social presence of the character is correlated with the level of anthropomorphism. However, the social presence of the morph was rated significantly lower than both the more and the less human-like version of the character. This could be related to the categorization ambiguity of the morph, which is often related to more uncanny feelings in the literature [34]. Further investigations, e.g., with agents that are rated more extreme on the scale of uncanny, as well as the investigation of other potentially influential distal cues are left for future work.

This work is based on the Brunswik lens, a simple but limited model that outlines how an observer attributes states to another entity. In order to further address open challenges in human-machine interaction, in future work there is a need to develop richer models also taking into account theories from social psychology and neuroscience.

## VII. Conclusions

In this paper we investigated the influence of embodiment features on a human observer when attributing emotional states to an artificial agent. We demonstrated that both agent's embodiment and level of human-likeness influence participant's ability to correctly identify emotional facial expressions. More specifically, uncanniness and social presence were identified as related traits that can confound the attribution process. Social traits of the observer, i.e., personality and level of empathy, only played a minor role in the attribution of emotional state, but heavily influenced the perception of social presence in the agent. With our work we contributed to demonstrate that, as in human-human interactions, the Brunswik lens model applies to the attribution process in a human observer when perceiving the emotional state of an artificial character. In addition, our work provided novel insights in the relation between 2D virtual agents and 3D blended embodiments by using a back-projected robot.

## References


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

[2] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *Int. Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[3] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R. J. van der Werf, and L.-P. Morency, "Virtual rapport," in *Int. Workshop on Intelligent Virtual Agents*, 2006, pp. 14–27.

[4] E. Brunswik, *Perception and the representative design of psychological experiments*. Univ of California Press, 1956.

[5] A. Vinciarelli and A. Pentland, "New social signals in a new interaction world: the next frontier for social signal processing," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 1, no. 2, pp. 10–17, 2015.

[6] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[7] D. V. Becker, D. T. Kenrick, S. L. Neuberg, K. Blackwell, and D. M. Smith, "The confounded nature of angry men and happy women," *J. of Personality and Social Psychology*, vol. 92, no. 2, p. 179, 2007.

[8] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[9] J. C. Hager, P. Ekman, and W. V. Friesen, "Facial Action Coding System," *Salt Lake City, UT: A Human Face*, 2002.

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

[11] R. Hortensius, F. Hekele, and E. S. Cross, "The perception of emotion in artificial agents," in *PsyArXiv*, 2017.

[12] S. H. Seo, D. Geiskkovitch, M. Nakane, C. King, and J. E. Young, "Poor thing! would you feel sorry for a simulated robot?: A comparison of empathy toward a physical and a simulated robot," in *ACM/IEEE Int. Conference on Human-Robot Interaction*, 2015, pp. 125–132.

[13] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive Behavioural Systems*, 2012, pp. 114–130.

[14] R. Gifford, "A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior." *J. of Personality and Social Psychology*, vol. 66, no. 2, p. 398, 1994.

[15] J. Gratch and S. Marsella, "Lessons from emotion psychology for the design of lifelike characters," *Applied Artificial Intelligence*, vol. 19, no. 3-4, pp. 215–233, 2005.

[16] L. Chaby, I. Hupont, M. Avril, V. Luherne-du Boullay, and M. Chetouani, "Gaze behavior consistency among older and younger adults when looking at emotional faces," *Frontiers in Psychology*, vol. 8, p. 548, 2017.

[17] N. Lazzeri, D. Mazzei, A. Greco, A. Rotesi, A. Lanatà, and D. E. De Rossi, "Can a humanoid face be expressive? a psychophysiological investigation," *Frontiers in Bioengineering and Biotechnology*, vol. 3, p. 64, 2015.

[18] K. M. Lee, Y. Jung, J. Kim, and S. R. Kim, "Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction," *Int. Journal of Human-Computer Studies*, vol. 64, no. 10, pp. 962–973, 2006.

[19] S. Thellman, A. Silvervarg, A. Gulz, and T. Ziemke, "Physical vs. virtual agent embodiment and effects on social interaction," in *Int. Conference on Intelligent Virtual Agents*, 2016, pp. 412–415.

[20] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "How anthropomorphism affects empathy toward robots," in *ACM/IEEE Int. Conference on Human Robot Interaction*, 2009, pp. 245–246.

[21] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.

[22] T. Canli and Z. Amin, "Neuroimaging of emotion and personality: Scientific evidence and ethical considerations," *Brain and Cognition*, vol. 50, no. 3, pp. 414–431, 2002.

[23] P. Chevalier, J.-C. Martin, B. Isableu, and A. Tapus, "Impact of Personality on the Recognition of Emotion Expressed via Human, Virtual, and Robotic Embodiments," in *IEEE RO-MAN*, 2015.

[24] C. Breazeal, "Emotion and sociable humanoid robots," *Int. Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, 2003.

[25] "FaceGen Modeller," http://facegen.com/, accessed: 2018-04-10.

[26] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *Int. Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[27] A. M. Rosenthal-von der Pütten and N. C. Krämer, "How design characteristics of robots determine evaluation and uncanny valley related responses," *Computers in Human Behavior*, vol. 36, pp. 422–439, 2014.

[28] N. Aifanti, C. Papachristou, and A. Delopoulos, "The MUG facial expression database," in *IEEE Int. Workshop on Image Analysis for Multimedia Interactive Services*, 2010, pp. 1–4.

[29] G. Skantze and S. Al Moubayed, "IrisTK: a Statechart-based Toolkit for Multi-party Face-to-face Interaction," in *Int. Conference on Multimodal Interaction*, 2012, pp. 69–76.

[30] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german," *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[31] M. H. Davis, "A multidimensional approach to individual differences in empathy," *JSAS Catalog of Selected Documents in Psychology*, 1980.

[32] C.-C. Ho and K. F. 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.

[33] C. Harms and F. Biocca, "Internal consistency and reliability of the networked minds measure of social presence," 2004.

[34] J. Kätsyri, K. Förger, M. Mäkäräinen, and T. Takala, "A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness," *Frontiers in Psychology*, vol. 6, 2015.