

1

Does the Goal Matter? Emotion Recognition Tasks Can Change the Social Value of Facial Mimicry towards Artificial Agents

Giulia Perugia^{1,2*}, Maike Paetzel-Prüsmann³, Isabelle Hupont^{3,4}, Giovanna Varni⁶, Mohamed Chetouani⁵, Christopher Peters⁷ and Ginevra Castellano²

 ¹ Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, Netherlands
 ² Uppsala Social Robotics Lab, Department of Information Technology, Uppsala University, Uppsala, Sweden
 ³ Computational Linguistics, Department of Linguistics, University of Potsdam, Potsdam, Germany

 ⁴ Joint Research Centre, European Commission, Seville, Spain.
 ⁵ Institut des Systèmes Intelligents et de Robotique, Sorbonne Université, Paris, CNRS UMR 7222, France
 ⁶LTCI, Télécom Paris, Institut Polytechnique de Paris, France
 ⁷KTH Royal Institute of Technology, Department of Computational Science Technology, Stockholm, Sweden

Correspondence*: Giulia Perugia Uppsala Social Robotics Lab, Uppsala University, Department of Information Technology, Box 337, 75105 Uppsala, Sweden giulia.perugia@it.uu.se

2 ABSTRACT

In this paper, we present a study aimed at understanding whether the embodiment and 3 humanlikeness of an artificial agent can affect people's spontaneous and instructed mimicry of 4 its facial expressions. The study followed a mixed experimental design and revolved around an 5 emotion recognition task. Participants were randomly assigned to one level of humanlikeness 6 (between-subject variable: humanlike, characterlike, or morph facial texture of the artificial agents) 7 and observed the facial expressions displayed by three artificial agents differing in embodiment 8 (within-subject variable: video-recorded robot, physical robot, and virtual agent) and a human 9 (control). To study both spontaneous and instructed facial mimicry, we divided the experimental 10 11 sessions into two phases. In the first phase, we asked participants to observe and recognize the emotions displayed by the agents. In the second phase, we asked them to look at the 12 agents' facial expressions, replicate their dynamics as closely as possible, and then identify 13 the observed emotions. In both cases, we assessed participants' facial expressions with an 14 automated Action Unit (AU) intensity detector. Contrary to our hypotheses, our results disclose 15 that the agent that was perceived as the least uncanny, and most anthropomorphic, likable, and 16 co-present, was the one spontaneously mimicked the least. Moreover, they show that instructed 17 facial mimicry negatively predicts spontaneous facial mimicry. Further exploratory analyses 18 revealed that spontaneous facial mimicry appeared when participants were less certain of the 19

emotion they recognized. Hence, we postulate that an emotion recognition goal can flip the social value of facial mimicry as it transforms a likable artificial agent into a distractor. Further work is

22 needed to corroborate this hypothesis. Nevertheless, our findings shed light on the functioning of

23 human-agent and human-robot mimicry in emotion recognition tasks and help us to unravel the

24 relationship between facial mimicry, liking, and rapport.

Keywords: Human-Robot Interaction, Human-Agent Interaction, Affective Computing, Facial Mimicry, Anthropomorphism, Uncanny
 Valley, Facial Action Coding System

1 INTRODUCTION

The success of artificial agents in areas like healthcare, personal assistance, and education highly depends 27 on whether people perceive them as likable and pleasant to interact with. In the lab, people's perceptions of 28 an artificial agent can be easily measured with questionnaires and interviews. In real-life settings, instead, 29 the artificial agent is on its own and the explicit evaluation of the interaction is not always feasible. In 30 fact, in these contexts, people might skip the proposed surveys or reply carelessly due to lack of time 31 and interest (Chung and Cakmak (2018)). A more promising approach in such contexts may be the use 32 of behavioral measures. While behavioral measures are in general extensively used in Human-Agent and 33 Human-Robot Interaction (HAI and HRI), they are seldom linked to people's self-reported perceptions (e.g., 34 likability and engagement, see Perugia et al. (2020a, 2021)). In this paper, we focus on facial mimicry - the 35 mirroring of another person's facial expressions (Hatfield et al. (1992)) - and aim to provide a fundamental 36 understanding of whether humans mimic the facial expressions of the six basic emotions displayed by 37 artificial agents and whether facial mimicry is linked to people's perception of artificial agents. To the 38 best of our knoweldge, no previous work in HAI and HRI has focused on this goal, hence this paper 39 constitutes one of the first attempts to understand whether facial mimicry of the six basic emotions is 40 related to people's perceptions of artificial agents. 41

42 From psychology, we know that facial mimicry increases with rapport (Tickle-Degnen and Rosenthal (1990); Hess et al. (1995)), but also appears in first acquaintances between individuals as a sign of liking 43 (Chartrand and Bargh (1999); Kulesza et al. (2015)). Studies on facial mimicry in HAI and HRI have so 44 far mostly focused on whether artificial agents are liked better when given the ability to mimic a human 45 interaction partner (Hoegen et al. (2018); Numata et al. (2020); Riek et al. (2009)). Hofree et al. (2014) 46 were among the few researchers who investigated whether human interaction partners mimic the facial 47 expressions of artificial agents as well. In their study, they disclosed that people's mimicry of an android's 48 facial expressions of anger and happiness is connected with their perceptions of the agent's humanlikeness 49 only when the android is co-present. In our study, we extend Hofree et al. (2014)'s work by (i) including 50 a wider spectrum of artificial agents, (ii) employing an overall less realistic humanoid robot that allows 51 for easy alteration of facial cues (i.e., Furhat), (iii) focusing on all six basic emotions, and (iv) using a 52 computer vision technique in lieu of Electromyography (EMG) to estimate people's facial mimicry. With 53 respect to EMG, computer vision is far less obtrusive and hence more viable for field use. 54

In this study, we involved 45 participants in an emotion recognition task with three artificial agents varying in embodiment (i.e., physical Furhat robot, video-recorded Furhat robot, and a virtual agent) and humanlikeness (i.e., humanlike, characterlike, and morph). Moreover, we also included a human control condition to understand differences between mimicry of artificial and natural agents. The emotion recognition task used in our experiment was divided into two phases. In the first phase, participants were asked to observe the facial expressions of the six basic emotions as expressed by the three artificial agents

and a video-recorded human (i.e., the control), and pick the correct one from a list. In the second phase, 61 62 instead, they were asked to observe the same facial expressions in re-shuffled order, mimic their temporal dynamics as closely as possible, and only afterwards recognize them. Based on Kulesza et al. (2015), 63 64 this latter phase was carried out under the pretense that intentional mimicry of facial expressions could 65 actually improve participants' emotion recognition. Participants' faces were video-recorded in both stages of the experiment and the activation of the action units (AU) corresponding to the six basic emotions was 66 67 determined through Hupont and Chetouani (2019)'s AU intensity detector. In the first part of the study, we 68 gauged which facial expressions were *spontaneously* mimicked by participants. In the second part of the 69 study, we focused instead on participants' *instructed* mimicry, and estimated how accurate participants were in replicating the temporal dynamics of the observed facial expressions. 70

71 The aim of this study is to understand (1) whether an artificial agent's embodiment and humanlikeness 72 can influence people's spontaneous and instructed facial mimicry (as suggested by Hofree et al. (2014) 73 and Mattheij et al. (2013, 2015)), (2) if spontaneous facial mimicry is related to people's perceptions of artificial agents, especially in terms of anthropomorphism, social presence, likability, and uncanniness 74 (perceptual dimensions expected to be influenced by the agent's level of humanlikeness), and (3) whether 75 76 there is a link between instructed and spontaneous facial mimicry. The overarching ambition of this work is to explore whether spontaneous facial mimicry can be used as an implicit, unconscious cue of liking 77 and rapport in HAI and HRI, and whether instructed facial mimicry can act as its proxy in settings where 78 79 spontaneous facial mimicry is difficult to gauge. Our work contributes to efforts paying the way towards unobtrusive automatic assessment of facial mimicry in interactions with artificial agents, hence facilitating 80 the measurement of liking and rapport through behavioral cues in the future. 81

2 RELATED WORK

82 Facial mimicry is the spontaneous imitation of another individual's facial expression without explicit instruction to do so (Hatfield et al. (1992)). Within the area of facial mimicry research, emotional mimicry 83 refers to the spontaneous mirroring of a facial expression with inherent emotional meaning, for instance, 84 wincing when observing others in pain (Bavelas et al. (1986)) or frowning at another person's frown. This 85 86 paper focuses on people's mimicry of the six basic emotions - happiness, sadness, surprise, anger, fear, 87 and disgust (Ekman and Rosenberg (1997)) - as displayed through the facial expressions of artificial and natural agents. Within the subsections 2.1, 2.2, and 2.3, we give an account of the different theories on 88 89 the nature and functioning of spontaneous facial mimicry in human-human interactions (HHI). We then describe the literature on human-agent and human-robot facial mimicry in subsection 2.4, and explain 90 91 our interest in instructed facial mimicry in subsection 2.5. Since the extant HAI and HRI literature has 92 hardly ever explored the relationship between facial mimicry and people's perceptions of artificial agents, in subsection 2.4 and 2.5, we mostly present studies focused on changes in facial mimicry due to an agent's 93 humanlikeness and co-presence. 94

95 2.1 Nature of Spontaneous Facial Mimicry

96 There are two main theoretical perspectives on the nature and functioning of emotional mimicry: a motor 97 and an emotional perspective. The *motor perspective* holds that emotional mimicry is an unconscious, 98 unintentional, unemotional, and reflex-like matching of observed facial expressions (Chartrand and Bargh 99 (1999)). Within this context, the *associative sequence learning* (ASL) approach posits that mimicry happens 98 by virtue of a learned long-term association between an action stimulus (e.g., a person's smile) and an 99 action response (e.g., the observer's smile; Heyes (2011)), which holds as long as the action stimulus (e.g., the observed facial expression) is similar to other stimuli previously associated with a certain motor action(e.g., the observer's facial expression).

104 Another theoretical formalization within the motor perspective is the *automatic embodiment account*, 105 which postulates that mimicry is the embodied motor simulation of an observed emotion that serves the 106 purpose of emotion recognition (Niedenthal et al. (2010)). According to this approach, we mimic another 107 individual's facial expressions to better recognize and differentiate them.

As opposed to the motor perspective, the *emotional perspective* sees mimicry as a marker of subtle 108 affective states arising in response to emotional stimuli (Dimberg (1990, 1997)). Within this perspective, 109 the facial-feedback hypothesis (Tomkins (1984); Izard (2013)), which dates back to Darwin (Darwin and 110 Prodger (1998)), posits that "the sight of a face that is happy, loving, angry, sad, or fearful (...) can cause 111 the viewer to mimic elements of that face and, consequently, to catch the other's emotions" (Hatfield et al. 112 (1992)). With a slightly different line of thought, the *affect-matching account* suggests that observing 113 a facial expression triggers a corresponding affective state in the observer, which then generates the 114 mimicking act (Dimberg et al. (2000)). Within the emotional perspective, there is hence no clear consensus 115 yet as to whether the affective state arising from an emotional stimulus precedes or succeeds mimicry. 116

The motor and emotional perspectives make somewhat different claims on the outcomes of emotional 117 mimicry (Moody et al. (2007). The motor perspective assumes that facial mimicry is always consistent 118 with the observed facial expression (i.e., emotion-congruent mimicry). For instance, an expression of 119 anger can only trigger a corresponding expression of anger. On the opposite, the emotional perspective 120 suggests that mimicry is related to the action tendencies associated with a stimulus (e.g., competitive and 121 collaborative tasks, Lanzetta and Englis (1989)). Thus an expression of anger can trigger anger but also fear 122 (i.e., valence-congruent mimicry), and the type of emotion triggered depends on the meaning associated 123 with the observed facial expression and the context where mimicry takes place (Fischer et al. (2012)). 124

125 2.2 Evidence Supporting Theoretical Accounts on Spontaneous Facial Mimicry

126 In general, there is little experimental support for the motor perspective. Available studies almost exclusively focused on facial mimicry of happiness and anger. As Hess and Fischer (2013, 2014) underline, 127 such studies only confirm that people display a valence-congruent facial expression when exposed to 128 happiness and anger (i.e., smiling to happiness, frowning to anger). However, they do not fully back up 129 emotion-congruent facial mimicry, which is at the core of the motor perspective. With regards to the 130 automatic embodiment account, several studies have investigated whether blocking facial mimicry impairs 131 the correct recognition of emotional facial expressions (Niedenthal et al. (2001); Hawk et al. (2012)). 132 Current evidence supports this position only partially. Indeed, mimicry seems to be crucial for emotion 133 recognition but only when it comes to recognizing ambiguous or subtle facial expressions (Hess and Blairy 134 (2001); Fischer et al. (2012)). 135

There are a number of studies that support the emotional perspective. For instance, Laird and Bresler 136 (1992) noticed that when people are asked to reproduce facial expressions of fear, anger, sadness, and 137 disgust, they also report experiencing those emotions. Moreover, Ekman et al. (1983) note that the muscular 138 reproduction of the facial expressions of the six basic emotions activates the Autonomic Nervous System 139 (ANS) in a similar way as to when people actually experience those emotions. Finally, Dimberg and 140 Thunberg (1998) describe how the facial response system that is responsible for mimicry responds to 141 emotions faster (300-400 ms) than the ANS (1-3 sec.), thus finding support for the affect-matching account. 142 Further support for the emotional perspective was also brought by Moody et al. (2007) who found that fear 143

priming elicits expressions of fear in response to both fear and anger, thus demonstrating that mimicry isnot a purely automatic mirroring of an observed emotion, but has an intrinsic emotional meaning.

146 2.3 The Social Value of Spontaneous Facial Mimicry

Regardless of their different views on the nature of facial mimicry, both the motor and the emotional 147 perspective posit that facial mimicry serves a social purpose. In one case (i.e., motor perspective), it serves 148 to recognize and respond to other people's emotions. In the other case (i.e., emotional perspective), it serves 149 the purpose of emotional contagion (Hatfield et al. (1993); Varni et al. (2017)), as to say "the tendency to 150 151 automatically mimic and synchronize movements, expressions, postures, and vocalizations with those of another person and, consequently, to converge emotionally" (Hatfield et al. (1992)). The literature suggests 152 that mimicry is indicative of higher liking during first acquaintances (Chartrand and Bargh (1999); Kulesza 153 154 et al. (2015); Calvo-Barajas et al. (2020)), stronger rapport in already established relationships (Hess et al. (1995)) and that it increases when two interaction partners are given the goal to affiliate (Lakin et al. (2003)). 155 In fact, Hess et al. (1995) found that watching funny movies with friends elicits more laughs than watching 156 them with strangers. Consistently, Fischer et al. (2012) discovered that dyads of friends mimic each other's 157 smiles of pride more than strangers do. Hess and Fischer (2013) and Bourgeois and Hess (2008) propose 158 that mimicry acts as a *social regulator* as it communicates the intention to bond. Since emotional mimicry 159 is known to be related with interpersonal stance (Prepin et al. (2013)), social tuning (Bernieri (1988)), 160 bonding (Jaques et al. (2016)), and rapport (Tickle-Degnen and Rosenthal (1990); Gratch et al. (2006); 161 Wang and Gratch (2009)), we consider it an important phenomenon to study in Human-Robot (HRI) and 162 Human-Agent Interaction (HAI). In fact, if facial mimicry was found to work similarly for artificial agents 163 164 and humans, it could inform future work investigating its use as an implicit and unconscious measure of 165 the quality of interaction in HAI and HRI (Perugia et al. (2020b)).

166 2.4 Spontaneous Facial Mimicry of Virtual Agents and Social Robots

In face-to-face interactions between humans, acted facial expressions constitute the only possibility of 167 168 studying spontaneous facial mimicry in a controlled way. However, acted facial expressions can be perceived by humans as being inauthentic and hence might hinder the occurrence of mimicry. For this reason, in 169 170 psychology, studies on spontaneous facial mimicry have almost exclusively focused on static images or 171 videos of facial expressions, with these latter being sometimes used to simulate live video-sessions (Kulesza et al. (2015)). With respect to humans, virtual and robotic agents give the unique possibility to investigate 172 173 spontaneous mimicry in face-to-face interactions occurring in real-time while preserving control over the 174 experimental setup (Hoegen et al. (2018)). This is because they enable researchers to manipulate only a few facial action units (AU) and control their activation over time. In this sense, the use of virtual and robotic 175 176 agents not only allows to investigate whether spontaneous facial mimicry occurs or not in specific contexts, but also opens up the possibility to understand whether its temporal dynamics are replicated. 177

178 While human-agent mimicry has been explored more thoroughly (Gratch et al. (2006); Hoegen et al. (2018)), studies on human-robot mimicry gained popularity more recently. Such a delay is probably due to 179 the fact that robots' faces were not provided with enough degrees of freedom to accurately reproduce facial 180 181 expressions until very recently. Most available studies on human-robot and human-agent mimicry focus on 182 endowing agents with the ability to mimic the facial expressions of human interactants and observing how this ability affects people's perceptions and reactions (Hoegen et al. (2018); Numata et al. (2020); Riek 183 184 et al. (2009)). Only a few studies investigate people's spontaneous mimicry of an artificial agent's facial 185 expressions, and, to the best of our knowledge, only one of these discussed the relationship between facial mimicry and people's perceptions of a robot (i.e., humanlikeness Hofree et al. (2014)). 186

The available studies in HAI and HRI show similar results to human-human mimicry, with the main 187 188 difference residing in the lower intensity and slower speed of human-agent and human-robot mimicry. For instance, Mattheij et al. (2015, 2013) found evidence for the spontaneous mimicry of happiness, surprise, 189 and disgust in the context of HAI and Philip et al. (2018) disclosed that people spontaneously mimic 190 virtual agents' facial expressions of joy, anger, and sadness. They also observed that mimicry is less intense 191 when it is directed to a virtual agent with respect to a human one. Similarly, in HRI, Hofree et al. (2014) 192 observed that people mimic a video-recorded android (i.e., Hanson's Einstein robot) to a lesser extent than a 193 video-recorded human. Furthermore, they discovered that, while the facial expressions of a video-recorded 194 195 android are mimicked only when the robot is perceived as highly humanlike, physically co-present androids are mimicked regardless of the perceptions they elicit. Hence, they proposed that it is the robot's co-196 presence that makes its humanlike appearance highly salient, and in turn elicits spontaneous facial mimicry. 197 Following this line of thought, in the present study, we manipulated the artificial agents' humanlikeness, 198 as well as their embodiment, and attempted to understand whether these influenced spontaneous facial 199 mimicry. We employed all three embodiments used by Hofree et al. (2014) - a video-recorded human, 200 a video-recorded robot, and a physical robot. Moreover, we added a virtual agent as in Mattheij et al. 201 (2013, 2015). In line with Li (2015), we considered: (1) the video-recorded robot as artificial, physically 202 embodied, but not co-present; (2) the physical robot as artificial, physically embodied, and co-present; 203 and (3) the video-recorded human as natural, physically embodied, but not co-present. While Li (2015) 204 differentiates between physical and digital co-presence, in this work we combined the two into one single 205 206 category of *co-presence* to distinguish between the two video-recordings that capture behavior of the past and hence do not share the same environment and time with the participant (i.e., video-recorded robot 207 and video-recorded human) from the virtual agent which shares the same environment and time with the 208 participant. Consequently, we categorize the virtual agent as artificial, virtually embodied, and co-present. 209

In HHI, Bourgeois and Hess (2008) showed that the social context in which the interaction takes place has 210 the power to influence emotional mimicry. While happy expressions are mimicked regardless of whether an 211 212 observed person is an in-group or out-group member, expressions of sadness are mimicked only between in-group members. Likewise, in HRI, Hofree et al. (2018) showed that participants mimicked a robot's 213 smiles and frowns when cooperating with it, but displayed inverse mimicry (i.e., frowned at the robot's 214 smiles and smiled in response to its frowns), when the context was competitive. To circumvent this problem, 215 in this study, we showed the agents' facial expressions to participants in a non-interactive context inspired 216 by Kulesza et al. (2015). Similar to Hofree et al. (2014), in this study, we asked participants to carefully 217 observe the agents' facial expressions. Inspired by Kulesza et al. (2015), however, we also gave them the 218 goal to recognize the emotion displayed by the agent. 219

220 2.5 Spontaneous and Instructed Facial Mimicry

Facial mimicry can further be divided into spontaneous and instructed. Spontaneous facial mimicry, which 221 we have discussed so far, occurs unconsciously, without any specific instruction (Hatfield et al. (1992)). 222 Instructed facial mimicry, instead, is deliberate mimicry of facial expressions that occurs consciously as a 223 result of specific instructions (McIntosh et al. (2006); Paetzel et al. (2017)). In their study, Hofree et al. 224 225 (2014) used instructed facial mimicry to ensure that the facial expressions of the android they used were 226 visible, feasible to imitate, and that electromyography (EMG) was working properly. Interestingly, they reported similar results for spontaneous and instructed facial mimicry. In fact, similar to spontaneous 227 facial mimicry, the instructed facial mimicry of the video-recorded android was less intense than the one 228 directed to the video-recorded human. This result brought us to hypothesize that instructed facial mimicry 229 might be somehow linked to spontaneous facial mimicry. To deepen our understanding of the relationship 230

between instructed and spontaneous facial mimicry, in this paper, we explore whether spontaneous facial mimicry can be predicted by people's ability to accurately reproduce the dynamics of an agent's facial expressions of the six basic emotions upon instruction to do so. Moreover, we study whether artificial agents' embodiment and level of humanlikeness can affect instructed facial mimicry in a way that is analogous to spontaneous facial mimicry. Should instructed facial mimicry be found to significantly predict spontaneous facial mimicry, it could be used as an *explicit* cue of people's social tuning with an artificial agent and could act as proxy of spontaneous facial mimicry.

3 RESEARCH QUESTIONS AND HYPOTHESES

In this work, we explore the influence of embodiment and humanlikeness on people's spontaneous and 238 instructed mimicry of artificial agents' facial expressions of the six basic emotions. Based on Hofree 239 et al. (2014) and Mattheij et al. (2013, 2015), we chose three artificial embodiments for this study: a 240 video-recorded robot, a physical robot, and a virtual agent. Furthermore, we added a control condition 241 in which participants observed the facial expressions of a video-recorded human. To change the artificial 242 agents' level of humanlikenss, we manipulated their facial features to resemble those of a characterlike 243 face, a humanlike face, and a face that includes features from both of them (i.e., a morph). Humanlikeness 244 was chosen as an independent variable in our study not only because Hofree et al. (2014)) found it to 245 be salient for facial mimicry, but also since it is known to influence people's perceptions of an agent's 246 anthropomorphism, social presence, and uncanniness (Mori et al. (2012), Kätsyri et al. (2015)), which are 247 perceptual dimensions that in turn affect liking and rapport. Our first group of research questions (RQ1a -248 RQ1c) concerns spontaneous facial mimicry: 249

- (RQ1a) To what extent does the humanlikeness of artificial agents influence people's spontaneous facial mimicry?
- (RQ1b) To what extent does the embodiment of artificial agents influence people's spontaneous facial mimicry?
- 254 (**RQ1c**) Does spontaneous facial mimicry differ between artificial and human agents?

Our second group of research questions (RQ2a - RQ2c) revolves around instructed facial mimicry. In previous work (Paetzel et al. (2017)), we investigated how well people were able to reproduce the dynamics of a laughter performed by an artificial agent that they were explicitly instructed to mimic. In this paper we focus on facial expressions of the six basic emotions instead. Here, we aim to understand whether the agents' embodiment and humanlikeness can affect instructed facial mimicry similar to how they affect spontaneous facial mimicry. Therefore, we pose the following research questions:

- (RQ2a) To what extent does the humanlikeness of artificial agents influence people's ability to mimic their facial expressions as accurately as possible when instructed to do so?
- (RQ2b) To what extent does the embodiment of artificial agents influence people's ability to mimic their facial expressions as accurately as possible when instructed to do so?
- 265 (**RQ2c**) Does instructed facial mimicry differ between artificial and human agents?

The long-term goal of our research is to provide initial insights on the development of implicit and explicit behavioral measures that can extend or replace questionnaire-based investigations of the perception of artificial agents. Previous work in human-human interaction has highlighted that spontaneous facial mimicry signals liking in first acquaintances (Chartrand and Bargh (1999); Kulesza et al. (2015)) and rapport in established relationships (Hess et al. (1995); Fischer et al. (2012)). Liking and rapport are complex constructs known to be influenced by factors such as the appearance and embodiment of an agent (Perugia et al. (2021); Paetzel et al. (2020); Paetzel-Prüsmann et al. (2021)). In this study, besides understanding the role of embodiment and humanlikeness in facial mimicry, we aim to gain more insights on the relationship between spontaneous facial mimicry and a few of the perceptual dimensions known to influence rapport and liking. The relationship between facial mimicry and people's perceptions of artificial agents has been addressed only seldom in the HAI and HRI literature.

(RQ3) To what extent can spontaneous facial mimicry predict the agent's perceived social
 presence, anthropomorphism, uncanniness, and likability?

From the related literature, we know that the occurrence of spontaneous facial mimicry can be an important predictor of the rapport people build with a human or artificial interaction partner. However, due to occlusions of the face and the subtlety of the mimicked facial expressions, it is often difficult to capture and quantify spontaneous facial mimicry in natural settings and more complex interactions. In these contexts, instructed facial mimicry could act as a proxy of spontaneous facial mimicry and could be used in place of a questionnaire as an *explicit* indirect cue of liking and rapport. Our fourth research question is thus concerned with the relation between instructed and spontaneous facial mimicry:

286 (**RQ4**) To what extent does instructed facial mimicry predict spontaneous facial mimicry?

Based on related studies performed by Hofree et al. (2014), Chartrand and Bargh (1999), Kulesza et al.
(2015), and Hess et al. (1995), we expected that:

(H1) Physically embodied, co-present, humanlike artificial agents elicit higher spontaneous 289 facial mimicry with respect to virtually embodied, non-co-present, non-humanlike 290 artificial agents. 291 (H2) Physically embodied, co-present, humanlike artificial agents elicit higher *instructed* 292 facial mimicry with respect to virtually embodied, non-co-present, non-humanlike 293 artificial agents. 294 (H3) Spontaneous facial mimicry positively predicts people's evaluations of the agents' 295 anthropomorphism, social presence, and likability, and negatively predicts their 296 297 perceived uncanniness. (H4) Instructed facial mimicry positively predicts spontaneous facial mimicry. 298

4 METHODOLOGY

299 Our study followed a 3x3+1 mixed experimental design with:

- *Embodiment* as within-subject variable with three types of artificial embodiment: a virtual agent, a
 physical Furhat robot (Al Moubayed et al. (2012)), and a video-recording of the Furhat robot (cf. Fig. 1)
- *Humanlikeness* as between-subject variable with three levels of humanlikeness: *humanlike*, *characterlike* and a *morph* between the humanlike and the characterlike (cf. Fig. 2)

Furthermore, we included a control condition (+1) for the embodiment in which participants observed a video-recorded human (cf. Fig. 1). This control condition was the same across all levels of the agent's humanlikeness.

The experimental design was informed by Kulesza et al. (2015) and consisted of two parts. In the first part, each participant was asked to observe the facial expressions of the agents and identify which of the



Figure 1. Different types of artificial embodiments used in the experiment. From left to right: a virtual agent; a physical Furhat robot; and a video recording of the Furhat robot.

six basic emotions they displayed (i.e., happiness, sadness, surprise, anger, fear, disgust). In the second
part, which occurred after a 5-minute break, participants were explicitly told that the accuracy of mimicry



Figure 2. Levels of humanlikeness used in the experiment in the Furhat robot (top) and virtual agent (bottom). Left: characterlike; right: humanlike; center: morph.

311 could improve emotion recognition. Consequently, they were instructed to observe the facial expressions 312 corresponding to the six basic emotions performed by the same agents (in randomized order), mimic them 313 as closely as possible, and identify them only after they finished mimicking. The first part of the experiment 314 allowed us to study spontaneous facial mimicry, the second part to investigate instructed facial mimicry. 315 Participants were video-recorded during both parts of the study.

Each participant observed a set of facial expressions performed by the video-recorded human and the 316 three artificial agents. All three artificial agents had the same level of humanlikeness but differed in 317 their embodiment. Each set of facial expressions was composed of expressions of the six basic emotions 318 performed twice by each agent. Within each set, the order of presentation of the stimuli was randomized, 319 and no two facial expressions of the same type occurred one after the other. The order of presentation of the 320 artificial and human agents was shuffled using Latin Squares. In total, each participant observed 48 facial 321 322 expressions for each part of the study. Emotional facial expressions were presented in short sequences of 5 seconds including onset, apex and offset, without vocalizations nor head movements (cf. Fig. 3). 323

324 4.1 Participants

An a priori sample size calculation for a repeated measures ANOVA with within-between interactions 325 (3x3), medium effect size (.25), α error probability of .05, and Power of 0.9 revealed a required total 326 sample size of 45. We hence recruited 46 participants from an international study program in Computer 327 Science at Uppsala University. Participants had at least a high school degree and came from a diverse 328 geographic background (44.4% Swedish). The 46 participants were randomly allocated to the three 329 conditions corresponding to the different levels of humanlikeness of the artificial agents: characterlike 330 (N=15; 11 male; 4 female; 0 other/prefer not to say), *humanlike* (N=16; 13 male; 3 female; 0 other/prefer 331 not to say), and morph (N=15; 12 male; 3 female; 0 other/prefer not to say). Due to a misunderstanding of 332 the study task, we excluded the data of one male participant from the humanlike condition. The final sample 333 of participants had a mean age of 26.16 years (SD= 4.37) and was composed of 10 people identifying 334 themselves a female and 35 as male. None of the participants had previously interacted with the Furhat 335 robot. 336

337 4.2 Embodiment and Humanlikeness

As a robot, we chose the Furhat platform (Al Moubayed et al. (2012)). Furhat is a blended robot head consisting of a rigid mask on which a facial texture is projected from within. We chose the Furhat robot for this experiment as its virtual face allowed us to easily alter facial features and design smooth and noiseless facial expressions.

342 We designed three different facial textures for the artificial agents. The humanlike face was created from pictures of a real human face using the FaceGen Modeller¹. The characterlike face was the standard Furhat 343 face with sketched "drawing-like" lips and eyebrows. Finally, the morph face was created by blending 344 the humanlike and characterlike skin textures in the Paint.NET digital photo editing package. The three 345 different textures we applied to the artificial agents were selected from a set of 28 faces tested in a pre-study 346 347 on Amazon Mechanical Turk (AMT). Since initial experiments with the Furhat robot found the face mask 348 without any projection to be perceived as male and dominant (Paetzel et al. (2016)), we limited the set of stimuli to male faces. The same texture we used for the Furhat robot was also utilized to create the virtual 349 agent's face. The video-recorded robot was obtained by recording the physical Furhat. For the human 350

¹ https://facegen.com/modeller.htm



Figure 3. Facial expressions of the six basic emotions. From left to right: happiness, sadness, surprise, fear, anger and disgust. From top to bottom: virtual agent, Furhat robot.

351 control condition, instead, we selected the video-recordings of a male person from the MUG database352 (Aifanti et al. (2010)).

353 4.3 Synthesis of Facial Expressions

354 The human in the MUG database was video-recorded while performing the facial expressions of the 355 six basic emotions following the Facial Action Coding System (FACS, Hager et al. (2002); Ekman et al. (1978)) and an onset-apex-offset temporal scheme. We designed the facial expressions of the artificial 356 357 agents by replicating the dynamics of the human video recording as closely as possible. Unfortunately, as 358 in Furhat's IrisTK animation system (Skantze and Al Moubayed (2012)), some facial Action Units (AUs) are combined and cannot be controlled separately, the facial expressions of the human and those of the 359 360 artificial agents slightly differed (cf. Fig. 3). An expert trained in the FACS ensured that the final set of 361 stimuli for the artificial agents was still following the FACS' guidelines.

We conducted an online preliminary study with 60 participants recruited on Amazon Mechanical Turk 362 to assess whether the naturalness, recognition rate, and intensity of the facial expressions of the six basic 363 364 emotions we designed was comparable across artificial agents. Crowd-workers were shown videos of the 365 facial expressions of the six basic emotions plus a neutral expression displayed by either the virtual agent, 366 the Furhat robot, or a human control (between-subjects). Our validation study revealed that the facial expressions of the two artificial agents were comparable, except for a difference in sadness. As expected 367 given the limitations of the IrisTK animation system, the expression of anger displayed by the human 368 control was rated as less intense compared to the one displayed by the artificial agents, and the expression 369 of disgust displayed by the human control was rated as more intense compared to the one displayed by the 370 artificial agents. With regards to naturalness, the expression of happiness was perceived as more natural for 371 the human control than for the artificial agents, and the human video received a higher rating in naturalness 372 373 than the videos of the artificial agents. We can hence conclude that, while some differences between the embodiments prevail, they mostly concern differences between the human control vs. the artificial 374



Figure 4. Left: the experimental setup. Right: A participant in the participant area.

agents. Since the human video was only used as control in our study, and most of the analyses involvedcomparisons between artificial agents, we did not think these differences could affect the results.

377 4.4 Experimental Setup

The experimental sessions took place in a private laboratory room at Uppsala University (cf. Fig. 4). To 378 grant a feeling of privacy and an even background for the video-recordings, the participant's area was 379 separated from the researcher's area by a blue curtain. Black curtains positioned behind the Furhat robot 380 (FR) and the screen displaying the other agents ensured a good visibility of the agents from the participants' 381 perspective. Uniform lighting for the recordings was guaranteed through a professional lighting system 382 (PLS) composed by two lamps. These were the only light sources in the experiment space. As both Furhat 383 and the screen displaying the agents were sources of light themselves, the dark environment ensured a 384 good visibility of the facial expressions. 385

The participant (PR) was sitting in the participants' area at a distance of about 100 cm from the Furhat 386 robot or the screen. This value falls in the personal space of the participants according to Hall (1969). 387 The agents were thus close enough to the participants to be properly seen, but not too close to elicit an 388 intimidating feeling. The agents were placed on a table at approximately 100 cm from the ground, which 389 was roughly at eye level for the majority of participants. The video-recorded and the virtual agents were 390 presented on a screen in portrait orientation. Their size was calibrated to match the size of Furhat's head. 391 All embodiments were controlled by a desktop computer (M1). An iPad (iP), placed on the table in front of 392 the participant, was used for answering the questionnaires. 393

394 4.5 Measures

395 4.5.1 Facial Recordings

To record participants' faces, we used two LOGITECH C920HD PRO webcams (WB) with a 800x600 396 resolution, operating at 30 fps. The webcams were placed on top of a tripod. One was positioned in front 397 of participants, at approximately 60 cm from them, and slightly on their side to not occlude the stimulus. 398 The second was positioned on the side of the participant (cf. Fig. 4). The webcams were connected to a 399 laptop (M2) which was used to start, stop, and control the video-recordings during the experiment. Each 400 webcam recorded the entire experimental session with the exclusion of the break between the spontaneous 401 and instructed mimicry trials. Hence, we obtained two video files per camera, participant and session. 402 The video-recordings of the frontal camera were used to assess participants' mimicry, those of the lateral 403 camera to capture the entire experimental scene. 404

405 4.5.2 Questionnaires

Throughout the experiment, four different questionnaires were used. Questionnaire Q1 consisted of a general demographic questionnaire (10 items), the short version of the Big Five personality traits (10 items, Rammstedt and John (2007)) and the Interpersonal Reactivity Index (IRI, 21 questions, excluded personal distress, Cronbach's α between .70 and .78 according to Davis et al. (1980)). This questionnaire gauged the empathy and personality traits of the participants, and hence was not used to answer this paper's research questions.

412 Questionnaire Q2 was shown to participants after every facial expression they observed to assess the 413 emotion they recognized in the stimulus. It was composed of the question "Which of these facial expressions 414 was just displayed?" with the six basic emotions, "neutral" and "I don't know" as response options, and the 415 question: "How certain are you of the selection you made in question 1?" with a three point Likert scale 416 using the labels: "Uncertain", "Neither nor", "Certain". The response options in the first question were 417 displayed in one of three pre-shuffled orders to prevent a bias towards the first item on the scale.

- Questionnaire (Q3) was shown after every embodiment in the first part of the experiment (i.e., spontaneous
 mimicry trial) to measure participants' perceptions of the agents on four dimensions:
- Anthropomorphism (5 items, 5-point Likert scale), sub-scale from the Godspeed questionnaire by Bartneck et al. (2009) (Cronbach's $\alpha = .91$ according to Ho and MacDorman (2010)).
- Social presence (8 items, 5-point Likert scale), excerpt from the social presence questionnaire developed by Harms and Biocca (2004). Sub-scales: co-presence (2 items, α = .84), Attentional Allocation (2 items, α = .81), Perceived Affective Understanding (2 items, α = .86), Perceived Emotional Interdependence (1 item, α = .85) and Perceived Behavioral Interdependence (1 item, α = .82).

Uncanniness and Likability (10 items, 5-point Likert scale), excerpt from Rosenthal-von der Pütten and
 Krämer (2014), sub-scales likability and perceived threat (Cronbach's α >= 0.82 for both sub-scales).

429 The order of questions and items remained the same across all embodiments.

At the end of the experimental session, the experimenter performed a semi-structured interview with the participant. The interview covered potential previous interactions with the Furhat robot, whether participants found aspects in the appearance of one of the characters particularly eerie, and if they had the impression that some of the facial expressions they observed were more difficult to trace back to a specific emotion. This interview was used to gather additional information about the experiment, and was not used to answer any research question present in this paper.

436 4.6 Procedure

After arriving to the lab, participants were informed about the experimental procedure, signed a consentform and answered Q1 on the iPad in front of them.

For the first part of the experiment (cf. Fig. 5 left), participants were asked to first watch the facial expressions displayed by the four agents, which always started and ended with a beep tone, and then indicate which emotions they corresponded to using the questionnaire Q2 displayed on the iPad. Participants were also explained that, once they finished completing Q2 on the iPad and after a pause of about 2 seconds, the agent would automatically display the next facial expression preceded and followed by another beep tone, and the same procedure would be repeated until they had observed all facial expressions.



Figure 5. A visualization of the experiment procedure. Note that the order of embodiments as well as the facial expressions were shuffled between the experimental parts and participants to control for the ordering effect. In total, participants saw 12 expressions (2 x 6 emotions) per embodiment in each part.

After participants observed all 12 expressions (2 trials x 6 emotions) for one embodiment, they rated their perception of the observed agent using questionnaire Q3 on the iPad. When necessary, the experimenter used this lapse of time to switch the physical robot with the screen. Once the participant finished responding to Q3, the stimuli for the subsequent embodiment were shown. Once participants responded to Q3 for the fourth and final agent of the spontaneous mimicry condition, they were given a five minute break and served refreshments.

For the second part of the experiment (cf. Fig. 5 right), participants were told that research suggests 451 that mimicry increases emotion recognition. Therefore, they were asked to perform the same task once 452 453 again, but this time by first mimicking the facial expression as accurately as possible and then noting down the emotion. The second part of the experiment followed the same procedure of the first part but 454 the embodiments were re-shuffled in order. As Q3 was omitted for the second part of the experiment, 455 participants had a shorter break between embodiments. Participants were shown each emotional expression 456 twice in each phase of the study to make the estimation of their facial mimicry more robust. They were 457 asked to recognize every emotion they were shown, so as to control for a potential learning effect due to 458 multiple exposures. 459

At the end of the session, the experimenter conducted the short semi-structured interview. This was followed by a debriefing in which the researcher explained the true nature and objective of the experiment. Participants were informed again that they could request the deletion of their data at any point in time.

5 MIMICRY PROCESSING

The strategy to segment the videos differed between spontaneous and instructed facial mimicry. In the first case, we were interested in understanding whether people mimicked the observed facial expressions or not, whereas in the second case, we were interested in understanding how accurate people were in mimicking the dynamics of the observed facial expressions. This difference in focus is motivated by the different expected magnitudes of spontaneous and instructed facial mimicry. While the former is a subtle response that does not necessarily follow the same dynamics of the expression observed, the latter was expected to be a much stronger and accurate response due to its explicitly imitative nature.



Figure 6. AU intensity detection pipeline. The white crosses represent the facial landmarks extracted from the face, while the dashed lines link the landmarks used for aligning each ROI.

470 For spontaneous facial mimicry, we annotated the frontal videos of the corresponding trial with the beginning and end of each stimulus in the ELAN 5.4. software. To do so, we used the audible beep tones 471 that marked the start and end of each facial expression of the agents. We then used the minutes obtained 472 from the annotation to automatically cut the original video into shorter snippets using ffmpeg². To properly 473 474 divide the instructed mimicry episodes, instead, we first manually identified the initial and final mimicry 475 frames for each stimulus by closely examining the participant's AU activation, and then we cut the original 476 video a second before and after these frames. This process ideally led to 96 individual video snippets per participant, 48 for spontaneous and 48 for instructed facial mimicry. 477

Once the data were segmented, we deployed an automatic AU intensity detector to recognize which muscles of the participants' face were activated in each video snippet of the spontaneous and instructed facial mimicry trials (cf. subsection 5.1). Then, in the case of spontaneous facial mimicry, we checked the AU time series to understand whether or not the target AU or combination of AUs amounting to each facial expression was active for a given lapse of time (cf. subsection 5.2). In the case of instructed facial mimicry, instead, we used the AU time series to perform a Cross-Recurrence Quantification Analysis (CRQA, Varni et al. (2017)) as detailed in Section 5.3.

485 5.1 Detection of AU Activation

The AU intensity detector used in this work is presented in Hupont and Chetouani (2019) and follows the pipeline shown in Fig. 6. In a first step, it segments the face of the person from the whole input image and extracts a set of facial landmarks. Face segmentation is carried out by means of the Viola and Jones (2004)' Haar Cascade algorithm. The landmarks (14 white crosses in Figure 6) are extracted with the Intraface library introduced by Xiong and Torre (2013). On the basis of the facial landmark positions, three rectangular facial Regions of Interest (ROIs) are then defined and features of Histogram Oriented Gradients (HOG, Dalal and Triggs (2005)) are computed for each one of them. The ROIs used in our pipeline are:

² https://ffmpeg.org/

- Frown ROI (used for AU4 model): This ROI is located around the inner eyebrow landmarks, which are also used for alignment purposes.
- Eyes ROI (AU1, AU2 and AU6): This ROI is made up of 8 patches located around the inner eyebrows,
 the middle eyebrows and the eye landmarks. ROI alignment is performed using inner eye corners. The
 final descriptor results from the concatenation of the 8 HOG descriptors.
- Mouth ROI (AU12, AU15, AU20, AU25 and AU26): This ROI is bounded by the nose center, the two
 lip corners and the lower lip. Alignment is done with respect to the lip corner positions.

500 Finally, the classification of each AU in terms of intensity is performed by an individually pre-trained 501 Support Vector Machine (SVM) model using its corresponding ROI features as input. The SVM models 502 were trained on the large-scale DISFA facial action database (Mavadati et al. (2013)). Each model detects 503 the activation of its corresponding AU in terms of six intensity categories, which are, according to Ekman's 504 taxonomy, (Hager et al. (2002)): "N" (neutral), "A" (trace), "B" (slight), "C" (marked), "D" (severe) and 505 "E" (maximum). The AU detector achieved an overall Intraclass Correlation Coefficient ICC(3,1) of 0.73, 506 which is within state-of-the-art performances in the task of AU intensity detection.

507 The AU intensity time series was low-pass filtered through a centered moving average filter with a 508 window size of 10 samples (33.3ms). This filtering was applied to both the spontaneous and instructed 509 facial mimicry time series. Moreover, the duration of time for which each AU was activated was also 510 computed. For instructed facial mimicry, the first and the last 30 samples corresponding to the 1 second 511 buffer left before and after the initial and final mimicry frames were removed in the final time series.

512 5.2 Processing of Spontaneous Facial Mimicry

To assess spontaneous facial mimicry, we divided the AU time series into two time intervals. The first time interval spanned from 0 to 1000 ms after stimulus onset and encompassed quick mimicry responses occurring at a subperceptual level, which Dimberg et al. (2000) call Rapid Facial Reactions (RFR). The second time interval ranged from 1000 to 5000 ms after stimulus onset and comprised facial mimicry responses occurring at a more conscious level, which we call Controlled Facial Reactions (CFR).

To consider a facial expression as mimicked at each time interval (RFR, CFR), we checked whether the 518 AU or combination of AUs corresponding to the target facial expression (cf. Table 1 based on Ekman et al. 519 (1978)) was active for at least 3 consecutive frames (100 ms). The activation was coded as 0 (not activated) 520 or 1 (activated) and the intensity of the activation was not considered for this analysis as we expected the 521 intensity of spontaneous facial mimicry to be low. We chose the threshold of 100 ms based on Ito et al. 522 523 (2004), who defined this as the shortest period of time a muscle can take to move. To perform the statistical analyses, we calculated the percentage of spontaneous facial mimicry for RFR and CFR. This value was 524 525 obtained *per embodiment* by dividing the number of trials in which the participant mimicked the facial expressions by the number of valid video snippets for that embodiment. Since in the spontaneous mimicry 526 part of the study, participants were not explicitly asked to mimic the facial expressions they observed, in 527 some snippets their faces were occluded, out of frame, or not recognizable by the AU intensity detector. 528 These snippets were excluded from the final analyses. If more than half of the snippets of a particular 529 embodiment were missing, we also excluded the other valid snippets from the a final analysis. Overall, this 530 led to the exclusion of a total of 465 snippets for RFR (22%) and 394 for CFR (18%), and left us with 1695 531 valid snippets for RFR, and 1766 for CFR. 532

Table 1. AUs or combination of AUs used to detect the spontaneous mimicry of the facial expressions of the six basic emotions (based on Ekman et al. (1978))

Emotion	Action Units (AUs)
Anger	AU4
Disgust	AU4 + AU25
Fear	AU20, AU1 + AU2 + AU4
Happiness	AU6, AU12, AU6 + AU12
Sadness	AU1, AU15, AU1 + AU4
Surprise	AU26, AU1 + AU2



Figure 7. Cross-Recurrence Plot of two different participants intentionally mimicking the facial expressions of anger displayed by the virtual agent (left) and the physical Furhat (right).

533 5.3 Analysis of Instructed Facial Mimicry

In order to accurately assess the dynamics of facial expressions, we performed a CRQA analysis (Marwan et al. (2007)). CRQA is a technique enabling a quantitative measure of the graphical patterns occurring in a Cross-Recurrence Plot (CRP, cf. Fig. 7). CRP is a plot looking at the times at which the features of a dynamical system *recur* (i.e., it is *close*) to features of another dynamical system. In this study, the two dynamical systems were the user and the artificial agents / video-recorded human, and the features were the AU intensities.

A CRP can be displayed as a square / rectangular black and white area spanned by two time series describing two systems. Black points correspond to the times at which the two systems co-visit the same area in the feature space, whereas white points correspond to the times at which each system runs in a different area. A CRP is expressed by the following *cross-recurrence matrix* (CR) :

$$CR_{i,j}^{\vec{f1},\vec{f2}}(\epsilon) = \Theta(\epsilon - \|\vec{f1}_i - \vec{f2}_j\|), \ i = 1...N, j = 1...M$$
(1)

where $\vec{f1}$ and $\vec{f2} \in \mathbb{R}^d$ are the d-dimensional time series of the two systems having N and M samples, respectively; ϵ is the threshold to claim closeness between two points, $\Theta(.)$ is the Heaviside function and $\|.\|$ is a norm. In this study, $\vec{f1}$ and $\vec{f2} \in \mathbb{R}^3$ are the time series of the AU intensities of the human and the artificial agents / video-recorded human over N samples. The threshold ϵ was set to 2 expressing that there was a match only when the 'distance' between the intensities of corresponding AUs was less than two. Thenorm used was the Manhattan distance.

550 CRPs can be analyzed through the Cross-Recurrence Quantification Analysis (CRQA) that enables to 551 extract quantitative information from the black and white patterns appearing in the plot (see Marwan et al. 552 (2007) for a complete survey). Typical patterns are: single isolated points, periodical diagonal lines, and 553 vertical / horizontal lines. These patterns are hints of randomness, periodicity and laminar states of the 554 dynamics of the system. In this study, we focused on the following CRQA measures (Marwan et al. (2007)):

555 *Cross-Recurrence Rate* (*cRR*)

556 The Cross-Recurrence Rate is defined as:

$$cRR(\epsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} CR_{i,j}(\epsilon)$$
⁽²⁾

557 and measures the density of recurrence points in a CRP. It corresponds to the ratio between the number 558 of the matrix elements shared by the participant and the artificial agents / video-recorded human and the 559 number of available elements (i.e. all the elements in the matrix). Here, cRR represents the overall extent 560 to which the human and the artificial agent / recorded human were activating the same AUs at a similar 561 level. This measure alone, however, even if it is a first measure to address mimicry, does not provide any 562 information about how mimicry unfolds over time. To extract information about that, three other CRQA 563 measures were computed:

564 Average diagonal lines length (L) and maximum diagonal line length (L_{max})

565 L represents the average length of a recurrent trajectory in a CRP. It is defined as:

$$L = \frac{\sum_{l=l_m}^{N} lP(l)}{\sum_{l=l_m}^{N} P(l)}$$
(3)

where l_m is the minimal diagonal length to be taken into account, and P(l) is the histogram of the diagonal lines. The minimal diagonal length was set to 8 samples, i.e. around 250 ms (Fasel and Luettin (2003)). The value of L expresses how stable a recurrent trajectory is. Here high values of L correspond to long, almost identical portions of AU intensities of the human and the artificial agents over time. Moreover, the length L_{max} of the longest diagonal line in the CRP was computed. A large value of L_{max} shows a slow divergence of the AUs' intensity trajectories.

572 Determinism (DET)

573 As a fourth and last measure, the determinism was computed. It is defined as:

$$DET = \frac{\sum_{l=l_m}^{N} lP(l)}{\sum_{l=1}^{N} P(l)}$$
(4)

574 It measures the percentage of the cross-recurrence points forming diagonal lines (of at least length l_m) 575 computed with respect to all the cross-recurrence points in the CRP. *DET* ranges in [0, 1] and it is a hint of 576 the predictability of the system (when DET = 0 the systems is stochastics, when DET = 1 it is periodic). 577 In this study, high values of DET were expected to be found during good mimicry episodes. 578 While participants paid more attention to stay in frame during the instructed mimicry phase, we still had 579 to exclude snippets due to occlusions and errors of the AU intensity detector. If there were only 5 or less 580 valid snippets for a particular embodiment and participant, these were removed from the final analysis. 581 Overall, we excluded a total of 209 snippets (9%) and were left with 1951 valid snippets for the analysis of 582 instructed mimicry. For the statistical analysis, we calculated the average cRR, L_{max} , L_{avg} and DET of 583 each participant across all valid trials associated with one embodiment.

6 **RESULTS**

In the remainder of the paper, we use: (1) social presence to refer to the dependent variables co-presence, 584 585 attentional allocation, perceived affective understanding, perceived emotional interdependence, and 586 perceived behavioral interdependence; (2) perception of the agent to refer to the dependent variables 587 anthropomorphism, likability, and perceived threat; (3) emotion recognition to refer to the dependent 588 variables percentage of correctly recognized emotions and average confidence in the recognized emotion; 589 (4) spontaneous facial mimicry to refer to the percentage of spontaneous facial mimicry for rapid facial 590 reactions (RFR) and controlled facial reactions (CFR); (5) instructed facial mimicry to refer to the average 591 (avg) cRR, avg L, avg L_{max} , and avg DET.

For the two manipulation checks (MC1 and MC2) and the preliminary analyses (PA), and for answering RQ1 and RQ2, we performed separate 3x3 mixed measures ANOVAs with humanlikeness as *betweensubject factor* (i.e., humanlike, characterlike, morph), embodiment as *within-subject factor* (i.e., virtual agent, physical robot, and video-recording of the physical robot) and (i) social presence (MC1), (ii) perception of the agent (MC2), (iii) emotion recognition (PA), (iv) spontaneous facial mimicry (RQ1) and (v) instructed facial mimicry (RQ2) as *dependent variables*. All p-values that we report in the post-hoc analyses are Bonferroni corrected to account for multiple tests.

599 For MC2, PA, RQ1, and RQ2, we also ran follow-up 2x3 mixed measures ANOVAs with humanlikeness 600 as a between-subject factor (humanlike, characterlike, morph), artificiality of the agent as a within-subject 601 factor (i.e., artificial agents and human agent), and the same dependent variables. To perform these analyses, 602 we calculated the average value across all three artificial agents on each dependent variable. Social presence 603 (MC1) was excluded from this set of analyses since the video-recorded human did not vary in embodiment 604 like the artificial agents. We kept humanlikeness as a between-subject factor to control for eventual effects 605 of the different levels of humanlikeness of the artificial agents on the dependent variables. However, as this 606 effect is already covered by the 3x3 mixed measures ANOVAs, for the sake of brevity, we do not report these results. All the p-values that we report in the post-hoc analyses are Bonferroni corrected to account 607 608 for multiple tests.

Finally, for answering RQ3 and RQ4, we performed separate regression analyses using spontaneous facial mimicry as a predictor of social presence and perceptions of the artificial agents (RQ3) and instructed facial mimicry as a predictor of spontaneous facial mimicry (RQ4). As RQ3 specifically focused on artificial agents' facial mimicry, we used only the data from the artificial agents to perform the regression analyses. On the contrary, as RQ4 focused on facial mimicry in general and not specifically on artificial agents' mimicry, we included also the data from the human video in the regression analyses.

615 6.1 Manipulation Check and Preliminary Analyses

616 6.1.1 Manipulation Check: Social Presence of the Artificial Agents

617 The results indicated a significant main effect of embodiment on co-presence, affective understanding, and 618 emotional interdependence (cf. Table 2 for the complete results). Furthermore, they showed a significant 619 interaction effect of humanlikeness and embodiment on co-presence.

Post-hoc analyses uncovered that the virtual agent was perceived as significantly more co-present than 620 the video-recorded robot (p = .005, cf. Table 3 for the descriptive statistics), and the physical robot was 621 perceived as significantly more co-present (p = .005) than the video-recorded one. No such difference 622 was observed between the virtual agent and the physical robot (p = 1.00). Moreover, they disclosed that 623 participants perceived their affective understanding of the physical robot to be significantly higher than that 624 of the virtual agent (p = .045), while the virtual agent and the video-recorded robot did not differ in terms 625 of perceived affective understanding (p = .255), and neither did the physical robot and the video-recorded 626 one (p = 1.00). Finally, participants perceived significantly higher emotional interdependence with the 627 physical robot with respect to both the virtual agent (p = .021, cf. Table 3 for the descriptive statistics) and 628 the video-recorded robot (p = .019). No such difference was present between the virtual agent and the 629 video-recorded robot (p = 1.00). 630

631 Further follow-up post-hoc analyses on the interaction effect of humanlikeness and embodiment on 632 co-presence uncovered that, in the characterlike condition, the virtual agent (M = 4.10, SD = .632) and the physical robot (M = 4.27, SD = .729) were perceived as significantly more co-present than 633 634 the video-recorded robot (M = 3.70, SD = .621, virtual agent: p = .026; physical robot: p = .028), but they did not significantly differ in co-presence from each other (p = 1.00). Likewise, in the morph 635 condition, the virtual agent (M = 3.71, SD = 1.051) was perceived as significantly more co-present than 636 the video-recorded robot (M = 3.07, SD = .938, p = .016), the physical robot (M = 3.53, SD = .930) 637 was perceived as significantly more co-present than the video-recorded one (p = .051), and the virtual agent 638 and the physical robot did not differ from each other (p = .409). Interestingly though, in the humanlike 639 condition (virtual agent: M = 3.79, SD = .777; physical robot: M = 4.04, SD = .746; video-recorded 640 robot: M = 4.00, SD = .734), these differences between artificial agents were not present (virtual agent -641 physical robot: p = .331; virtual agent - video-recorded robot: p = .083; physical robot - video-recorded 642 robot: p = 1.00). 643

644 6.1.2 Manipulation Check: Perception of the Agents

When checking for differences in the perception of the artificial agents across levels of humanlikeness and embodiments, we found a significant main effect of embodiment on anthropomorphism, perceived threat, and likability (cf. Table 2 for the complete results) and a significant main effect of humanlikeness on anthropomorphism.

Bonferroni-corrected post-hoc analyses revealed that the virtual agent and the video-recorded robot were 649 perceived as less anthropomorphic than the physical robot (both p < .001, cf. Table 3 for the descriptive 650 statistics), but the virtual agent and the video-recorded robot did not differ in terms of anthropomorphism 651 between each other (p < .682). Similarly, in terms of likability, the virtual agent (p = .001, cf. Table 3 for 652 the descriptive statistics) and the video-recorded robot (p = .042) were perceived as less likable than 653 the physical robot. However, the virtual agent and the video-recorded robot did not differ from each other 654 (p = .291). Finally, concerning perceived threat, the virtual agent was perceived as more threatening than 655 the video-recorded robot (p = .001, cf. Table 3 for the descriptive statistics), but no such difference was 656

	Em	bodiment	t	Huma	Humanlikeness			Embod. x Human.		
Social Presence	F(2, 80)	p	ηp^2	F(2, 40)	<i>p</i>	ηp^2	F(4, 80)	p	ηp^2	
Co-presence	7.878	.001	.165	2.719	.078	.120	4.036	.005	.168	
Att. Allocation	2.040	.137	.049	.036	.965	.002	1.377	.249	.064	
Aff. Understand.	3.643	.031	.083	1.115	.338	.053	2.373	.059	.106	
Em. Interdep.	5.864	.004	.128	2.157	.129	.097	.668	.616	.032	
Beha. Interdep.	.630	.535	.015	.750	.479	.036	.725	.578	.035	
Agent's Percept.	F(2, 80)	p	ηp^2	F(2, 40)	<i>p</i>	ηp^2	F(4, 80)	p	ηp^2	
Anthropomorph.	15.587	< .001	.280	3.399	.043	.145	2.246	.071	.101	
Perceived Threat	6.470	.002	.139	.244	.785	.012	2.447	.053	.109	
Likability	8.361	.001	.173	1.776	.182	.082	1.454	.224	.068	
Emo. Recogn.	F(2, 80)	p	ηp^2	F(2, 40)	p	ηp^2	F(4, 80)	p	ηp^2	
Recogn. (Spont.)	.296	.745	.007	4.004	.026	.167	.253	.907	.013	
Recogn. (Instr.)	1.147	.323	.028	5.540	.008	.217	.482	.749	.024	
Spont. Mimicry	F(2, 64)	p	ηp^2	F(2, 32)	p	ηp^2	F(4, 64)	p	ηp^2	
Freq. RFR	9.336	<.001	.226	1.002	.378	.059	1.071	.378	.063	
Freq. CFR	4.645	.013	.127	.636	.536	.038	1.566	.194	.089	
Instr. Mimicry	F(2,76)	p	ηp^2	F(2, 38)	<i>p</i>	ηp^2	F(4, 76)	p	ηp^2	
Avg cRR	.097	.908	.003	2.189	.126	.103	.785	.538	.040	
Avg L	.364	.696	.009	.208	.813	.011	.411	.800	.021	
Avg L _{max}	.477	.662	.012	.293	.748	.015	.298	.878	.015	
Avg DET	.219	.804	.006	.187	.830	.010	.784	.539	.040	

Table 2. Results of 3x3 Mixed Measures ANOVAs for Manipulation checks and RQ1 and RQ2. The significant results are displayed in bold, while the trend effects are presented in italics.

657 present between the virtual agent and the physical robot (p = .104) and between the video-recorded and 658 the physical robot (p = .822)

With regards to the main effect of humanlikeness, the post-hoc analyses disclosed that humalike artificial agents were perceived as significantly more anthropomorphic than morph artificial agents (p = .046, cf. Table 4 for the descriptive statistics). However, humanlike and characterlike artificial agents (p = .232) and characterlike and morph agents (p = 1.00) did not differ significantly from each other.

663 When running the 2x3 ANOVA focusing on the agents' artificiality, we found out that the video-recorded 664 human was perceived as significantly more anthropomorphic (p < .001), more likable (p < .001), and less 665 threatening (p < .001) than the artificial agents (cf. Table 5 for the results and the descriptive statistics).

666 Discussion of Manipulation Check

667 As specified in section 2.4, the artificial agents and the video-recorded human differed as follows: (i) 668 the physical robot was *artificial*, *physically embodied*, and *co-present*; (ii) the virtual agent was *artificial*, 669 *virtually embodied*, and *co-present*; (iii) the video-recorded robot was *artificial*, *physically embodied*,

	Virtual Agent		Physic	al Robot	Video Robot		
	M	SD	M	SD	M	SD	
Co-presence	3.87	.832	3.95	.844	3.59	.847	
Att. Allocation	4.05	.837	4.20	.757	4.02	.809	
Aff. Understanding	3.17	.778	3.45	.625	3.27	.658	
Em. Interdependence	1.79	.888	2.12	1.051	1.70	.832	
Beha. Interdependence	2.28	1.076	2.35	1.066	2.21	.914	
Anthropomorphism	2.53	.834	3.10	.742	2.67	.777	
Perceived Threat	2.10	.769	1.84	.650	1.72	.524	
Likability	2.22	.759	2.66	.708	2.43	.787	
Recognized (Spont.)	.77	.139	.79	.139	.77	.129	
Recognized (Instr.)	.80	.151	.77	.142	.79	.117	
Freq. RFR	.56	.165	.53	.189	.66	.151	
Freq. CFR	.74	.174	.66	.188	.73	.156	
Avg cRR	13.12	7.430	13.54	6.692	13.26	6.585	
Avg L	5.52	2.987	5.29	2.754	5.66	1.656	
Avg L _{max}	6.59	3.596	6.36	3.364	6.88	3.321	
Avg DET	2.78	1.856	2.63	1.612	2.78	1.262	

 Table 3. Descriptive Statistics of the 3x3 Mixed Measures ANOVAs per Embodiment: Mean (M) and standard deviation (SD) of all dependent variables

670 but not co-present; and (iv) the video-recorded human was natural, physically embodied, but not copresent. The manipulation checks that we performed were aligned with these differences. Indeed, the 671 video-recorded robot was perceived as significantly less co-present than the virtual agent and physical 672 673 robot. Furthermore, the artificial agent that was physically embodied and co-present (i.e., the physical robot) was perceived as easier to understand affectively, more anthropomorphic, more likable, and elicited 674 more emotional understanding than the other artificial agents. Finally, the human agent was perceived as 675 676 more anthropomorphic, more likable, and less threatening than the artificial agents. As a result, we can state that the manipulation of embodiment worked as expected in this study. 677

With regards to the manipulation of humanlikeness, the core dependent variable that we expected to 678 change was anthropomorphism. The characterlike and morph robot did not differ in anthropomorphism and 679 neither did the characterlike and humanlike robot. However, in line with our expectations, the humanlike 680 robot was perceived as more anthropomorphic than the morph robot. As a result, we considered the 681 manipulation of humanlikeness only partially successful. This is surprising given that the characterlike 682 version has a clearly sketched appearance compared to the humanlike version, which is derived from a 683 human picture. It is interesting to note, however, that all three versions had a comparable high rating of 684 humanlikeness, which could potentially be explained by the very humanlike appearance of the Furhat robot 685 platform itself. Both the humanlike and the characterlike facial texture may have hence elicited a congruent 686 and overall humanlike perception. The morph, on the contrary, may have received the lowest rating of 687 humanlikeness due to the incongruence of the facial features. Even though this did not lead to an uncanny 688 feeling in participants, it could have still decreased its anthropomorphism. 689

	Character.		Huma	anlike	Morph		
	M	SD	M	SD	M	SD	
Co-presence	4.02	.720	3.94	.722	3.44	.722	
Att. Allocation	4.10	.734	4.12	.737	4.05	.737	
Aff. Understanding	3.19	.550	3.48	.550	3.24	.550	
Em. Interdependence	2.20	.775	1.64	.775	1.74	.775	
Beha. Interdependence	2.16	.910	2.52	.909	2.17	.909	
Recognized (Spont.)	.83	.089	.74	.090	.76	.090	
Recognized (Instr.)	.85	.108	.72	.109	.79	.109	
Anthropomorphism	2.69	.631	3.11	.632	2.51	.632	
Perceived Threat	1.87	.500	1.96	.501	1.83	.501	
Likability	2.40	.620	2.67	.621	2.23	.621	
Freq. RFR	.56	.137	.54	.135	.62	.137	
Freq. CFR	.68	.144	.70	.144	.75	.144	
Avg cRR	15.75	5.863	11.07	5.863	12.95	5.863	
Avg L	5.53	2.312	5.17	2.311	5.74	2.312	
Avg L _{max}	6.82	2.862	6.11	2.859	6.87	2.862	
Avg DET	2.60	1.310	2.91	1.309	2.70	1.310	

Table 4. Descriptive Statistics of the 3x3 Mixed Measures ANOVAs per level of humanlikeness: Mean (M) and standard deviation (SD) of all dependent variables

690 With regards to the manipulation of humanlikeness, it was also very interesting to discover that, when 691 the appearance of the artificial agents was humanlike, the differences in co-presence between the different 692 embodiments ceased to exist. This result seems to suggest that the humanlike appearance has in itself a 693 quality of co-presence that goes beyond the physical instantiation of an artificial agent.

694 6.2 Preliminary Analyses: Emotion Recognition

As a preliminary analysis, we checked whether participants' ability to recognize the emotions displayed by 695 the artificial agents differed across embodiments and levels of humanlikeness. Interestingly, we discovered 696 a main effect of humanlikess on the percentage of emotion recognized (cf. Table 2 for the complete results). 697 According to the results, participants' emotion recognition was better when participants observed the 698 699 characterlike agents with respect to when they observed the humanlike agents (cf. Table 4 for the descriptive statistics). This was true both in the spontaneous mimicry (p = .029) and in the instructed mimicry trials 700 (p = .006, cf. Fig. 8 top row). No such differences in emotion recognition were observed between 701 702 characterlike and morph agents (spontaneous mimicry trial: p = .154; instructed mimicry trial: p = .466) and between morph and humanlike agents across trials (spontaneous mimicry trial: p = 1.00; instructed 703 mimicry trial: p = .218). When it comes to the 2x3 ANOVAs focusing on the agents' artificiality, we 704 found a significant difference between artificial agents and the video-recorded human in terms of emotion 705 recognition only for the instructed mimicry trial (cf. Table 5 for the results and descriptive statistics). In 706 this case, the percentage of emotions correctly recognized was higher for the human with respect to the 707 artificial agents. 708

	Art	tificiality		Artif.	Agents	Human Video		
Agent's Percept.	F(1, 41)	p	ηp^2	M	SD	M	SD	
Anthropomorphism	130.064	<.001	.760	2.77	.659	4.13	.761	
Perceived Threat	29.800	<.001	.421	1.93	.544	1.51	.451	
Likability	39.159	<.001	.489	2.43	.626	3.00	.757	
Emo. Recogn.	F(2, 40)	p	ηp^2	M	SD	M	SD	
Recognized (Spont.)	1.470	.233	.035	.78	.096	.75	.151	
Recognized (Instr.)	7.494	.009	.158	.79	.119	.84	.120	
Spont. Mimicry	F(1, 32)	p	ηp^2	M	SD	M	SD	
Freq. RFR	34.835	<.001	.521	.60	.146	.84	.127	
Freq. CFR	4.323	.046	.119	.74	.144	.79	.156	
Instr. Mimicry	F(1, 39)	p	ηp^2	M	SD	M	SD	
Avg cRR	.653	.424	.016	13.92	7.172	13.29	10.124	
Avg L	.039	.844	.001	5.56	2.288	5.53	2.697	
Avg L _{max}	.206	.652	.005	6.70	2.824	6.55	3.369	
Avg DET	.249	.621	.006	2.75	1.270	2.65	1.638	

Table 5. Results of 2x3 ANOVAs and Descriptive Statistics. The significant results are displayed in bold. The Mean (M) and standard deviation (SD) of all dependent variables are divided per Artificial and Human Agents.

Discussion of Preliminary Analyses. As predicted, the facial expressions of the video-recorded human were easier to recognize in comparison to the facial expressions of the artificial agents. However, somewhat unexpectedly, and partially in conflict with this result, the facial expressions of the characterlike artificial agents were easier to recognize with respect to those of the humanlike artificial agents both for the spontaneous and instructed mimicry trials. We ascribe this results to the stylized appearance of the characterlike agents, which might have made their expressions more readable and recognizable than those of the other agents.

716 6.3 Results for Research Questions

717 6.3.1 Influence of Embodiment and Humanlikeness on Spontaneous Facial Mimicry [RQ1]

Results disclosed a significant main effect of embodiment on spontaneous facial mimicry both for
RFR and CFR (cf. Table 2 for the complete results). However, we did not find any significant effect of
humanlikeness and embodiment and humanlikeness alone on spontaneous facial mimicry.

Post-hoc analyses with a Bonferroni correction disclosed that for the RFR the video-recorded robot was 721 722 mimicked significantly more than the virtual agent (p = .003, cf. Table 3 for the descriptive statistics and cf. Fig. 8 bottom row for the boxplot) and physical robot (p = .001), and that the physical robot and 723 the virtual agent did not differ in spontaneous facial mimicry from each other (p = 1.00). With regards 724 to CFR, the post-hoc analyses showed that the physical robot was mimicked significantly less than the 725 video-recorded robot (p = .038, cf. Fig. 8 bottom row for the boxplot), while the virtual agent and the 726 video-recorded robot did not differ in terms of spontaneous facial mimicry (p = 1.00) and only a trend 727 728 difference was present between the virtual agent and the physical robot (p = .068).



Figure 8. Top: Boxplots showing the effect of level of humanlikeness on the percentage of emotions correctly recognized for the spontaneous and instructed mimicry trials. Bottom: Boxplots showing the effect of the agent's embodiment on frequency of spontaneous facial mimicry for RFR and CFR.

When taking into account the artificiality of the agent as the within-subject factor by analyzing the average mimicry frequency across all artificial agents in comparison to the mimicry frequency when observing the video recording of the human, we found a significant main effect of artificiality on spontaneous facial mimicry (cf. Table 5 for the results and the descriptive statistics). In this case, the video-recorded human was spontaneously mimicked significantly more than the artificial agent both for RFR and CFR.

Discussion of RO1. These results are somewhat complementary to those we found for the manipulation 734 checks. Indeed, it seems that the agent that elicited the highest ratings of co-presence, affective 735 736 understanding, emotional interdependence, anthropomorphism, and likability, namely the physical robot, was also the agent that was spontaneously mimicked the least. If we take the facial mimicry-rapport 737 hypothesis into account, this result is somewhat counterintuitive. Indeed, in line with this hypothesis, the 738 robot eliciting the most favorable relational ratings should have been the one spontaneously mimicked 739 the most. However, if we take the emotion recognition task into account, we can partially explain this 740 result. Recognizing the emotions of another agent is an activity that implies putting some distance between 741 the agent we observe and ourselves. It somewhat entails considering the agent we observe as a stimulus, 742 rather than a relational agent. In this sense, we can hypothesize that an agent that is perceived as more 743 socially present and elicits more positive perceptions is less good as a stimulus, it is more likely to act as a 744 distractor, and hence can hinder the goal of the emotion recognition task. It is interesting to note that, when 745

the agent is a human, this dynamic does not take place and the human is, as foreseeable, spontaneously mimicked more than the artificial agents. We can ascribe this result to the familiarity of the human stimulus. Indeed, the video-recorded human is undoubtedly more positively evaluated than the artificial agents and hence more likely to act as a distractor. However, it is also the stimulus with which participants are the most familiar and whose facial expressions they are more used to recognize.

751 6.3.2 Influence of Embodiment and Humanlikeness on Instructed Facial Mimicry [RQ2]

The results of the 3x3 mixed measures ANOVAs did not show any significant effect of embodiment and humanlikeness on instructed facial mimicry (cf. Table 2 for the complete results). Similarly, the results of the 2x3 mixed measures ANOVA did not disclose any significant effect of the agents' artificiality on participants' instructed facial mimicry (cf. Table 5 for the results and the descriptive statistics).

756 **Discussion of RQ2.** As opposed to Hofree et al. (2014), in our study, the results of instructed facial mimicry are not congruent with those of spontaneous facial mimicry. Based on our findings, we can state 757 that when facial mimicry is explicitly prompted, the agent's appearance and embodiment cease to have 758 759 an influence on it. This might be due to the fact that, when facial mimicry transforms itself into a purely 760 imitative act, it loses its social value, and hence those variables that would have likely affected it due to their relational value, such as the agents' level of humanlikeness and their embodiment, do not influence it 761 anymore. This assumption is further reinforced by the fact that people's ability to mimic an agent as closely 762 as possible does not differ also when artificial and human agents are taken into account. 763

6.3.3 Spontaneous Facial Mimicry as Predictor of Perceived Social Presence and Perceptions of Artificial Agents [RQ3]

The results of the regression analyses in Table 6 show that spontaneous facial mimicry for RFR was a negative predictor of co-presence, attentional allocation, and affective understanding, whereas spontaneous facial mimicry for CFR was a negative predictor of attentional allocation and emotional interdependence. Moreover, they showed that spontaneous facial mimicry for RFR was a negative predictor of people's perceptions of the artificial agents' likability and anthropomorphism, and spontaneous facial mimicry for CFR was a negative predictor of participants' perception of the agents' likability (cf. Table 6 for the complete results).

Discussion of RQ3. These results are in line with those of RQ1 and seem to suggest that, in this study, 773 the more the artificial agents were spontaneously mimicked, the less positive perceptions they elicited, the 774 less socially co-present they were perceived, and the less people felt emotionally connected with them 775 and capable of understanding their affective states. We assume that this result, which goes against most 776 777 of the literature focusing on the social function of spontaneous facial mimicry, can be ascribed to the emotion recognition task in which participants were involved. Our hypothesis is that, within an emotion 778 recognition task, spontaneous facial mimicry does not fulfill anymore a social function, but rather serves the 779 purpose of emotion recognition. In this context, the embodiment people rated as the most anthropomorphic, 780 likable and co-present (i.e., the physically present robot), was the one that people could relate to the most. 781 Consequently, they might have had an easier time understanding its facial expressions, and hence less 782 necessity to spontaneously mimic them. 783

To verify this assumption, we performed a few additional regression analyses using spontaneous facial mimicry for RFR and CFR as predictors and the percentage of correctly recognized facial expressions and the confidence in the recognized emotion as dependent variables. As we supposed, participants' spontaneous facial mimicry was a significant negative predictor of their certainty of the correctness of the

	Spont. Mimicry (% RFR)				Spont. Mimicry (% CFR			
Dependent Variables	β	t(149)	p	r^2	β	t(154)	p	r^2
Co-presence	183	-2.271	.025	.034	100	-1.246	.215	.010
Att. Allocation	211	-2.626	.010	.045	230	-2.920	.004	.053
Aff. Understanding	203	-2.527	.013	.041	127	-1.587	.115	.016
Em. Interdependence	077	940	.349	.006	204	2.581	.011	.042
Beha. Interdependence	076	932	.353	.006	098	-1.224	.223	010
Anthropomorphism	168	-2.074	.040	.028	112	-1.388	.167	.012
Perceived Threat	.055	.666	.506	.003	.124	1.546	.124	.015
Likability	191	-2.373	.019	.037	262	-3.359	.001	.069
% Recognized	169	-1.821	.071	.029	073	783	.435	.005
Confidence Recogn.	231	-2.520	.013	.053	121	-1.310	.193	.015

Table 6. Regression Analyses [RQ3]. Frequency of spontaneous facial mimicry as predictor of Social Presence and Perceptions of Artificial agents. The significant results are displayed in bold, while the trend effects are presented in italics.

recognized emotion and a trend negative predictor of their emotion recognition performance (cf. Table 6 for the complete results). This indicates that *the more participants spontaneously mimicked the artificial agents, the less they were confident in the emotion they recognized*. Such a result is particularly important as it corroborates the theory that facial mimicry serves the purpose of emotion recognition, but only when the emotions to recognize are ambiguous (Hess and Blairy (2001); Fischer et al. (2012)).

793 6.3.4 Instructed Facial Mimicry as Predictor of Spontaneous Facial Mimicry [RQ4]

The results of the regression analyses displayed in Table 7 show that the average cRR, L, L_{max} , and DET are all significant negative predictors of spontaneous facial mimicry for RFR but they do not equally predict spontaneous facial mimicry for CFR.

Discussion RO4 This result is extremely interesting as it suggests that, in this study, the more closely 797 participants mimicked the facial expressions of the agents when instructed to do so, the less likely they 798 799 were to spontaneously mimic the agents at an unconscious level of processing. Since we have seen that spontaneous facial mimicry for RFR was a negative predictor of participants' confidence in the recognized 800 801 emotion (and partially also of their ability to recognize the target emotion), it does not surprise that people that mimic an emotion well under instruction, actually mimic it less at a subperceptual level. Indeed, if 802 people are better able to mimic all the temporal dynamics of a target facial expression, they might also be 803 more capable of recognizing that target emotion. In this sense, as opposed to spontaneous facial mimicry, 804 instructed facial mimicry might signal a better understanding of the emotion. This finding entails that, even 805 though in an emotion recognition task, instructed facial mimicry does not behave similarly to spontaneous 806 facial mimicry, it still maintains a relation with it. 807

7 GENERAL DISCUSSION

This study investigated how the humanlikeness and embodiment of an artificial agent could influence people's mimicry of its facial expressions. Based on Hofree et al. (2014), we expected that physically

	Spon	t. Mimicr	Spont	. Mimici	ry (% (CFR)		
Predictors	β	t(141)	p	r^2	β	t(146)	p	r^2
Avg cRR	229	-2.783	.006	.052	093	-1.124	.263	.009
Avg L	186	-2.241	.027	.035	068	819	.414	.005
Avg L _{max}	175	-2.107	.037	.031	057	683	.496	.003
Avg DET	212	-2.567	.011	.045	098	-1.189	.236	.010

 Table 7. Regression Analyses [RQ4]. Features of instructed facial mimicry as predictors of the frequency of Spontaneous facial mimicry. The significant results are displayed in bold.

embodied, co-present, and humanlike artificial agents could elicit higher spontaneous and instructed facial 810 mimicry than virtually embodied, non-co-present, and less humanlike ones, and that instructed facial 811 mimicry could positively predict spontaneous facial mimicry. Moreover, based on the link between facial 812 mimicry and rapport, we postulated that spontaneous facial mimicry could positively predict participants' 813 evaluations of the agents' anthropomorphism, social presence, and likability, and negatively predict 814 their perceived uncanniness. Although our manipulation of embodiment was successful and the one of 815 humanlikeness partially successful, and the task we chose was taken from the existing literature (Hofree 816 et al. (2014) and Kulesza et al. (2015)), the results we obtained did not meet our expectations (cf. H1-H4 817 in section 3). We found that: (i) the physically embodied, co-present artificial agent (i.e., the physical 818 robot) was the one that was spontaneously mimicked the least regardless of its humanlikeness (cf. H1); 819 (ii) instructed facial mimicry did not behave congruently to spontaneous facial mimicry (cf. H2); (iii) 820 spontaneous facial mimicry negatively predicted anthropomorphism, social presence, and likability, and 821 did not predict uncanniness (cf. H3); and (iv) instructed facial mimicry negatively predicted spontaneous 822 facial mimicry (cf. H4). 823

While these results were surprising, their consistency led to a hypothesis that some element of the task 824 that was given to the participants hindered the social value of facial mimicry. Following the *automatic* 825 embodiment account (Niedenthal et al. (2010)), we postulated that the task's focus on emotion recognition 826 could have caused a change in the meaning of facial mimicry. Additional analyses confirmed our suspicion. 827 In fact, they indicated that the spontaneous facial mimicry of the artificial agents was a significant negative 828 predictor of participants' confidence in the emotion recognized. This result seems to suggest that, in the 829 context of human-agent and human-robot mimicry, the emotion recognition goal of a task can flip the social 830 value of spontaneous facial mimicry, and transform a physically embodied, co-present artificial agent into 831 *a distractor*. This may have arisen by chance due to elements of the study design and deserves further 832 exploration and replication. Due to the only partially successful manipulation of the robot's perceived 833 humanlikeness, it also requires further investigation whether this effect is really independent of the robot's 834 level of anthropomorphism. The primary objective of this study was to understand whether spontaneous 835 facial mimicry could be used as a cue of liking and rapport in HAI and HRI, and whether instructed 836 facial mimicry could act as a proxy of spontaneous facial mimicry. Although our findings do not meet our 837 expectations, the fact that they went in the exact opposite direction to our original hypotheses may suggest 838 that, in an emotion recognition task, spontaneous facial mimicry can still be used as a predictor of liking 839 and rapport, and instructed facial mimicry could still function as a predictor of spontaneous facial mimicry, 840 but they need to be envisioned as negative predictors rather than positive ones. Additional work is needed 841 to corroborate these preliminary results, and understand whether context alone (emotion recognition task vs. 842 social interaction) can influence the value of facial mimicry in HAI and HRI in the way we have described. 843

8 LIMITATIONS & FUTURE WORK

844 One limitation of the current experimental design is the focus on one particular robotic embodiment (i.e., the Furhat robot). While this platform has several advantages, like the easy alteration of facial features 845 846 and expressions, it is sometimes difficult to discern facial detail clearly. By keeping the robot platform 847 consistent across conditions, we could limit the influence of confounding factors on our results. However, this in turn reduced the strength of the manipulation of humanlikeness and could be the reason why we did 848 not see the agents' anthropomorphism differ between the characterlike and humanlike, and the characterlike 849 850 and morph conditions. Future studies should hence investigate facial mimicry in emotion recognition tasks 851 carried out with multiple humanoid robots differing in their embodiment and degree of realism to check whether our findings still hold. We also suggest to replicate our study involving a larger and more diverse 852 853 set of participants, particularly when it comes to academic background and gender.

854 Unlike most related work (e.g., Hofree et al. (2014)), in our experiment, we included stimuli covering 855 all six basic emotions (Ekman et al. (1978)). For the analyses of facial mimicry, however, we combined people's responses to the different emotions together and calculated an average facial mimicry value. 856 857 Especially given the differences observed between individual expressions in our validation study, it is fair to 858 assume that, while comprehensive, our results might not fit all six basic emotions equally. Another element of variation that one might need to control when studying facial mimicry is the observer's belief that an 859 860 agent's facial expression reflects its subjective emotional state. In future facial mimicry studies, it would be 861 interesting to include additional questionnaires capturing people's belief about the agent's emotional state when performing facial expressions, and their own emotional state before and after the experiment. While 862 863 a mixed-measure experimental design has its advantages especially in very controlled environments that 864 make recruiting a large group of participants difficult, we acknowledge that the repeated measures in the current study design was probably tiring for participants and may have lead to them paying less attention 865 to the stimuli they encountered as last. While we randomized the order of presentation of the different 866 867 conditions to control for any order effect, it would overall be beneficial to replicate this experiment with a 868 between-subject design., especially if potentially expanding the set of questionnaires to be included in the study. 869

Since our study was task-based, non-interactive, and devoid on an emotional context, the acted nature 870 871 of the agents' facial expressions was particularly clear. Future work should focus on bringing the study 872 of facial mimicry into more interactive and social contexts and assess whether facial mimicry could be 873 used in place of questionnaires to assess people's social attunement with artificial agents. An important pre-874 condition for using facial mimicry as a behavioral indicator of people's relationship with a robot is a robust 875 and non-intrusive assessment technique of people's facial expressions. While the computer-vision-based 876 approach discussed in this paper has shown promising results, further improvements are necessary to make it more robust with respect to different angles and light conditions. This is especially important if we want 877 to bring the study of facial mimicry to less controlled scenarios. 878

9 CONCLUSION

879 In the study presented in this paper, we involved participants in an emotion recognition task carried out with 880 artificial agents differing in their embodiment and degree of humanlikeness. In the first phase of the study, 881 we asked participants to observe the artificial agents' facial expressions and attempt to identify the emotions 882 they displayed. In the second phase of the study, instead, we asked participants to observe the agents' facial 883 expressions, mimic them as closely as possible, and then identify them. We used the first part of the study

to investigate the frequency of participants' spontaneous facial mimicry, and the second part to investigate 884 885 the accuracy of their instructed facial mimicry. The aim was to understand whether spontaneous mimicry of artificial agents' facial expressions can be used as a behavioral cue of liking and rapport, and whether 886 887 instructed facial mimicry could act as a proxy of its spontaneous counterpart. Our results suggest that, in an emotion recognition task, the physical instantiation of an artificial agent, together with its likability 888 and anthropomorphism, intrudes rather than promotes people's spontaneous facial mimicry. Furthermore, 889 results suggest that instructed facial mimicry negatively predicts spontaneous facial mimicry. Since the 890 participants in this study mimicked the facial expressions of the artificial agents more when they were 891 uncertain about the emotion to recognize, one possibility is that, in emotion recognition contexts, facial 892 mimicry serves the purpose of emotion recognition. Even though our results did not support our initial 893 894 hypotheses, they nevertheless show that spontaneous mimicry can be a behavioral cue of liking and rapport, and instructed facial mimicry a proxy of spontaneous facial mimicry. 895

CONFLICT OF INTEREST STATEMENT

896 The authors declare that the research was conducted in the absence of any commercial or financial897 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

GP wrote the paper, formulated the research questions, segmented the stimuli for spontaneous mimicry, 898 devised the methodology to analyze spontaneous mimicry, performed all the statistical analyses, created 899 the tables and graphs, and interpreted the results. MPP collaborated in writing the paper, conceived and 900 designed the study, wrote the program for the robot interaction, collected the data, segmented the stimuli 901 902 for instructed mimicry, post-processed the spontaneous and instructed mimicry data, and helped with the interpretation of the results. IH conceived and designed the study, synthesized the facial expressions of 903 the artificial agents, developed and ran the AU intensity detector, contributed to writing section 5, and 904 905 helped with the processing of spontaneous and instructed mimicry. GV conceived and designed the study, developed and ran the CRQA for the instructed mimicry, contributed to writing section 5, and helped with 906 907 the processing of instructed mimicry. MC, CP and GC gave advice on the design of the study, stimuli, read 908 and gave comments on the final version of the paper.

FUNDING

909 This work is partly supported by the Swedish Foundation for Strategic Research under the COIN project 910 (RIT15-0133). The work received funding from ROMEO2 and Labex SMART (ANR-11-LABX-65) 911 supported by French state funds managed by the ANR within the Investissements d'Avenir programme 912 under reference ANR-11-IDEX-0004-02. The authors are solely responsible for the content of this 913 publication.

ACKNOWLEDGMENTS

914 Thanks to A. Axelsson, R. Cort, A. Y. Gao, D. Golay, K. von Hausswolff, A. Jansson, M. Lind, I. Löscher,
915 G. Nauwerck and A. Persson for input on the experimental design of the experiment. I. Hupont partly
916 undertook this work before arriving to European Commission's Joint Research Centre (JRC, Seville), but
917 finalised in the JRC and in line with the JRC work program. She participated in the experimental part of

the study when working at Institut des Systèmes Intelligents et de Robotique (Sorbonne Université, Paris)and in the writing of the paper after joining JRC.

DATA AVAILABILITY STATEMENT

920 The dataset presented in this article is not readily available because it contains personal information on the

921 participants. An anonymized version of the data supporting the conclusions of this article can be made 922 available by the authors upon request.

REFERENCES

- Aifanti, N., Papachristou, C., and Delopoulos, A. (2010). The MUG facial expression database. In *IEEE Int. Workshop on Image Analysis for Multimedia Interactive Services*. 1–4
- Al Moubayed, S., Beskow, J., Skantze, G., and Granström, B. (2012). Furhat: a back-projected human-like
 robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems*. 114–130
- Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the
 anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int.*

929 Journal of Social Robotics 1, 71–81

- Bavelas, J. B., Black, A., Lemery, C. R., and Mullett, J. (1986). "i show how you feel": Motor mimicry as
 a communicative act. *Journal of personality and social psychology* 50, 322
- Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal behavior* 12, 120–138
- Bourgeois, P. and Hess, U. (2008). The impact of social context on mimicry. *Biological Psychology* 77, 343–352
- Calvo-Barajas, N., Perugia, G., and Castellano, G. (2020). The effects of robot's facial expressions on
 children's first impressions of trustworthiness. In 2020 29th IEEE International Conference on Robot
 and Human Interactive Communication (RO-MAN) (IEEE), 165–171
- Chartrand, T. L. and Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social
 interaction. *Journal of personality and social psychology* 76, 893
- 941 Chung, M. J.-Y. and Cakmak, M. (2018). "how was your stay?": Exploring the use of robots for gathering
 942 customer feedback in the hospitality industry. In 2018 27th IEEE International Symposium on Robot
 943 and Human Interactive Communication (RO-MAN) (IEEE), 947–954
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. vol. 1, 886–893
- Darwin, C. and Prodger, P. (1998). *The expression of the emotions in man and animals* (Oxford University
 Press, USA)
- Davis, M. H. et al. (1980). A multidimensional approach to individual differences in empathy. JSAS *Catalog of Selected Documents in Psychology*
- Dimberg, U. (1990). For distinguished early career contribution to psychophysiology: award address, 1988:
 facial electromyography and emotional reactions. *psychophysiology* 27, 481–494
- 952 Dimberg, U. (1997). Facial reactions: Rapidly evoked emotional responses. Journal of Psychophysiology
- Dimberg, U. and Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian journal of psychology* 39, 39–45
- Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial
 expressions. *Psychological science* 11, 86–89

- Ekman, P., Friesen, W. V., and Hager, J. C. (1978). *Facial action coding system: Investigator's guide*(Consulting Psychologists Press)
- Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes
 among emotions. *science* 221, 1208–1210
- 961 Ekman, P. and Rosenberg, E. L. (1997). What the face reveals: Basic and applied studies of spontaneous
 962 expression using the Facial Action Coding System (FACS) (Oxford University Press, USA)
- Fasel, B. and Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition* 36, 259–275
- Fischer, A., Becker, D., and Veenstra, L. (2012). Emotional mimicry in social context: the case of disgust
 and pride. *Frontiers in psychology* 3, 475
- Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., et al. (2006).
 Virtual rapport. In *Int. Workshop on Intelligent Virtual Agents*. 14–27
- Hager, J. C., Ekman, P., and Friesen, W. V. (2002). Facial Action Coding System. Salt Lake City, UT: A *Human Face*
- 971 Hall, E. T. (1969). *The Hidden Dimension*, vol. 1990 (Anchor Books New York)
- Harms, C. and Biocca, F. (2004). Internal consistency and reliability of the networked minds measure ofsocial presence
- Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1992). Primitive emotional contagion. *Review of personality and social psychology* 14, 151–177
- 976 Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1993). Emotional contagion. *Current directions in* 977 *psychological science* 2, 96–100
- Hawk, S. T., Fischer, A. H., and Van Kleef, G. A. (2012). Face the noise: Embodied responses to nonverbal
 vocalizations of discrete emotions. *Journal of personality and social psychology* 102, 796
- Hess, U., Banse, R., and Kappas, A. (1995). The intensity of facial expression is determined by underlying
 affective state and social situation. *Journal of personality and social psychology* 69, 280
- Hess, U. and Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial
 expressions and their influence on decoding accuracy. *International journal of psychophysiology* 40,
 129–141
- Hess, U. and Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review* 17, 142–157
- Hess, U. and Fischer, A. (2014). Emotional mimicry: Why and when we mimic emotions. *Social and Personality Psychology Compass* 8, 45–57
- 989 Heyes, C. (2011). Automatic imitation. Psychological bulletin 137, 463
- Ho, C.-C. and MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating
 an alternative to the godspeed indices. *Computers in Human Behavior* 26, 1508–1518
- Hoegen, R., Van Der Schalk, J., Lucas, G., and Gratch, J. (2018). The impact of agent facial mimicry
 on social behavior in a prisoner's dilemma. In *Proceedings of the 18th International Conference on*
- 994 Intelligent Virtual Agents. 275–280
- Hofree, G., Ruvolo, P., Bartlett, M. S., and Winkielman, P. (2014). Bridging the mechanical and the human
 mind: spontaneous mimicry of a physically present android. *PloS one* 9
- Hofree, G., Ruvolo, P., Reinert, A., Bartlett, M. S., and Winkielman, P. (2018). Behind the robot's smiles
 and frowns: In social context, people do not mirror android's expressions but react to their informational
 value. *Frontiers in neurorobotics* 12, 14
- 1000 Hupont, I. and Chetouani, M. (2019). Region-based facial representation for real-time action units intensity
- detection across datasets. *Pattern Analysis and Applications* 22, 477–489

- Ito, T., Murano, E. Z., and Gomi, H. (2004). Fast force-generation dynamics of human articulatory muscles.
 Journal of applied physiology 96, 2318–2324
- 1004 Izard, C. E. (2013). *Human emotions* (Springer Science & Business Media)
- Jaques, N., McDuff, D., Kim, Y. L., and Picard, R. (2016). Understanding and predicting bonding in
 conversations using thin slices of facial expressions and body language. In *International Conference on Intelligent Virtual Agents* (Springer), 64–74
- 1008 Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (2015). A review of empirical evidence on different
 1009 uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness.
 1010 *Frontiers in Psychology* 6
- Kulesza, W. M., Cisłak, A., Vallacher, R. R., Nowak, A., Czekiel, M., and Bedynska, S. (2015). The face
 of the chameleon: The experience of facial mimicry for the mimicker and the mimickee. *The Journal of Social Psychology* 155, 590–604
- 1014 Laird, J. D. and Bresler, C. (1992). The process of emotional experience: A self-perception theory.
- 1015 Lakin, J. L., Jefferis, V. E., Cheng, C. M., and Chartrand, T. L. (2003). The chameleon effect as social glue:
- Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior* 27, 145–162
- Lanzetta, J. T. and Englis, B. G. (1989). Expectations of cooperation and competition and their effects on
 observers' vicarious emotional responses. *Journal of personality and social psychology* 56, 543
- Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent
 robots, telepresent robots and virtual agents. *Int. Journal of Human-Computer Studies* 77, 23–37
- Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex
 systems. *Physics Reports* 438, 237 329
- 1024 Mattheij, R., Nilsenova, M., and Postma, E. (2013). Vocal and facial imitation of humans interacting
- with virtual agents. In 2013 Humaine Association Conference on Affective Computing and Intelligent
 Interaction (IEEE), 815–820
- Mattheij, R., Postma-Nilsenová, M., and Postma, E. (2015). Mirror mirror on the wall. *Journal of Ambient Intelligence and Smart Environments* 7, 121–132
- Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., and Cohn, J. F. (2013). DISFA: A spontaneous
 facial action intensity database. *IEEE Transactions on Affective Computing* 4, 151–160
- McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., and Wilbarger, J. L. (2006). When the social
 mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism.
 Developmental science 9, 295–302
- Moody, E. J., McIntosh, D. N., Mann, L. J., and Weisser, K. R. (2007). More than mere mimicry? the
 influence of emotion on rapid facial reactions to faces. *Emotion* 7, 447
- Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 98–100
- Niedenthal, P. M., Brauer, M., Halberstadt, J. B., and Innes-Ker, Å. H. (2001). When did her smile drop?
 facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition & Emotion* 15, 853–864
- Niedenthal, P. M., Mermillod, M., Maringer, M., and Hess, U. (2010). The simulation of smiles (sims)
 model: Embodied simulation and the meaning of facial expression. *Behavioral and brain sciences* 33,
 417
- Numata, T., Sato, H., Asa, Y., Koike, T., Miyata, K., Nakagawa, E., et al. (2020). Achieving affective human-virtual agent communication by enabling virtual agents to imitate positive expressions. *Scientific reports* 10, 1–11
 - 040 *reports* 10, 1–1

- 1047 Paetzel, M., Perugia, G., and Castellano, G. (2020). The persistence of first impressions: The effect of repeated interactions on the perception of a social robot. In Proceedings of the 2020 ACM/IEEE 1048 1049 International Conference on Human-Robot Interaction. 73–82 Paetzel, M., Peters, C., Nyström, I., and Castellano, G. (2016). Congruency matters - How ambiguous 1050 gender cues increase a robot's uncanniness. In International Conference on Social Robotics 1051 Paetzel, M., Varni, G., Hupont, I., Chetouani, M., Peters, C., and Castellano, G. (2017). Investigating the 1052 influence of embodiment on facial mimicry in hri using computer vision-based measures. In 2017 26th 1053 IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 1054 1055 579–586 Paetzel-Prüsmann, M., Perugia, G., and Castellano, G. (2021). The influence of robot personality on the 1056 development of uncanny feelings. Computers in Human Behavior 120, 106756 1057 Perugia, G., Díaz-Boladeras, M., Català-Mallofré, A., Barakova, E. I., and Rauterberg, M. (2020a). Engage-1058 dem: a model of engagement of people with dementia. *IEEE Transactions on Affective Computing* 1059 Perugia, G., Paetzel, M., and Castellano, G. (2020b). On the role of personality and empathy in human-1060 human, human-agent, and human-robot mimicry. In International Conference on Social Robotics 1061 (Springer), 120–131 1062 Perugia, G., Paetzel-Prüsmann, M., Alanenpää, M., and Castellano, G. (2021). I can see it in your 1063 eyes: Gaze as an implicit cue of uncanniness and task performance in repeated interactions with robots. 1064 Frontiers in Robotics and AI 8, 78 1065 1066 Philip, L., Martin, J.-C., and Clavel, C. (2018). Rapid facial reactions in response to facial expressions of emotion displayed by real versus virtual faces. *i-Perception* 9, 2041669518786527 1067 Prepin, K., Ochs, M., and Pelachaud, C. (2013). Beyond backchannels: co-construction of dyadic stancce 1068 by reciprocal reinforcement of smiles between virtual agents. In Proceedings of the Annual Meeting of 1069 the Cognitive Science Society. vol. 35 1070 Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short 1071 version of the big five inventory in english and german. Journal of Research in Personality 41, 203-212 1072 1073 Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., and Robinson, P. (2009). How anthropomorphism affects empathy toward robots. In ACM/IEEE Int. Conference on Human Robot Interaction. 245-246 1074 Rosenthal-von der Pütten, A. M. and Krämer, N. C. (2014). How design characteristics of robots determine 1075 1076 evaluation and uncanny valley related responses. Computers in Human Behavior 36, 422-439 1077 Skantze, G. and Al Moubayed, S. (2012). IrisTK: a Statechart-based Toolkit for Multi-party Face-to-face Interaction. In Int. Conference on Multimodal Interaction. 69–76 1078 Tickle-Degnen, L. and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. 1079
- 1079 Tickle-Degnen, L. and Kosentnal, K. (1990). The nature of rapport and its nonverbal correlates. 1080 *Psychological inquiry* 1, 285–293
- 1081 Tomkins, S. S. (1984). Affect theory. Approaches to emotion 163
- Varni, G., Hupont, I., Clavel, C., and Chetouani, M. (2017). Computational study of primitive emotional
 contagion in dyadic interactions. *IEEE Transactions on Affective Computing* 11, 258–271
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision* 57, 137–154
- Wang, N. and Gratch, J. (2009). Rapport and facial expression. In 2009 3rd International Conference on
 Affective Computing and Intelligent Interaction and Workshops (IEEE), 1–6
- 1088 Xiong, X. and Torre, F. (2013). Supervised descent method and its applications to face alignment. In IEEE
- 1089 Conference on Computer Vision and Pattern Recognition. 532–539