How an Android Expresses "Now Loading…": Examining the Properties of Thinking Faces

Shushi Namba[1,2] ⬤ · Wataru Sato[2] · Saori Namba[1,2] · Alexander Diel[3] · Carlos Ishi[4] · Takashi Minato[4]

Abstract

The "thinking face" is a facial signal used to convey being in thought. For androids, the thinking face may be important to achieve natural human–robot interaction. However, the facial pattern necessary for portraying the thinking face remains unclear and has not yet been investigated in androids. The current study aims to (a) identify the facial patterns when people are engaged in answering complex questions (i.e., thinking face) and (b) clarify whether implementing the observed thinking faces in an android can facilitate natural human–robot interaction. In Study 1, we analyze the facial movements of 40 participants after they are prompted with difficult questions and indicate five facial patterns that corresponded to thinking faces. In Study 2, we further focus on the pattern of furrowing of the brows and narrowing of the eyes among the observed thinking facial patterns and implement this pattern in an android. The results show that thinking faces enhance the perception of being in thought, genuineness, human-likeness, and appropriateness in androids while decreasing eeriness. The free-description data also revealed that negative emotions are attributed to the thinking face. In Study 3, we compared the thinking vs. neutral faces in a question–answer situation. The results showed that the android's thinking face facilitated the perception of being in thought and human-likeness. These findings suggest that the thinking face of androids can facilitate natural human–robot interaction.

Keywords Thinking face · Human–robot interaction · Android · Dynamic facial expressions · Facial signals

1 Introduction

Facial signals have multiple functions in communication, and there has been increasing interest in their pragmatic use [1, 2]. For example, facial signals play the role of emphasizing a word or function as a back channel [3]. Investigation of facial signals can lead to several practical advancements. For example, research on facial signals can help realize more human-like interactions with android robots.

In this study, we investigated one of the important facial signals, the "thinking face," which is used to convey being in thought. We aimed to (a) identify the facial patterns when humans are engaged in answering complex questions (i.e., the thinking face) and (b) clarify whether implementing the observed thinking faces in an android can facilitate natural human–robot interaction.

1.1 Related Work

1.1.1 Human–Robot Interaction

Here, we define the term "android (robot)," based on the relevance and scope of this paper as a robot with a human-like face that can move its facial components in a human-like manner. Communication with others contributes to well-being [4, 5]; in light of this, the field of human–robot interaction aims to achieve more natural communication between robots and humans [6]. In particular, it has been indicated that most robots for natural user interaction have faces because facial expressiveness is considered a key component for developing personal attachment, along with prosodic expressiveness [7].

✉ Shushi Namba
nashushi@hiroshima-u.ac.jp

1 Department of Psychology, Hiroshima University, Hiroshima, Japan

2 RIKEN, Psychological Process Research Team, Guardian Robot Project, Kyoto, Japan

3 School of Psychology, Cardiff University, Cardiff, UK

4 RIKEN, Interactive Robot Research Team, Guardian Robot Project, Kyoto, Japan

Regarding android robots, there have been several studies on human–robot interaction in which robots express emotions and communicate such information [8, 9]. One of the reasons for the development of this research is that research on facial expressions has presented forms that correspond to emotions [10]. For example, Lazzeri et al. [9] showed that there was a trend that robots can express emotions better than emotional expressions through 2D photos or 3D models (see review Stock-Homburg [11]). Considering a more natural human–robot interaction, a common scenario involves a person posing a question to an android robot. In such situations, it would be deemed preferable for the non-human agent to exhibit a face that looks like it is thinking about the question, as opposed to a neutral face without facial movements.

1.1.2 Thinking Face

However, compared to facial expressions of emotion, the precise properties of facial signals that convey "being in thought" remain unclear. One well-known form, the "thinking face," has been studied in sociology and communication research [12]. Notably, scholars have investigated the "thinking face" not so much as a specific facial pattern but rather as a description focused on how it functions in particular contexts. The descriptions of the morphological properties of the "thinking face" have varied among previous research. For example, the thinking face has been described as having an averted gaze [12], looking upward with raised eyebrows [1], raising or furrowing eyebrows, closing eyes, pulling part of the mouth [1, 3], or wandering eyes [13]. More recently, Nota et al. [14] used a rich dialogue corpus to analyze the facial signals that occur during conversation and demonstrated that blinking, gaze shifting, raising eyebrows, and smiling occur during responses to questions. In sum, the movements around the eyes, including the eyebrows and gaze, can be considered important components of the thinking face.

To date, few studies have empirically investigated the production of the thinking face. Bitti et al. [15] investigated facial expressions that convey "I am thinking about it." The results showed that raising of the eyebrows, frowning of the eyebrows, and narrowing of the eyelids occurred when deliberately conveying a thinking face. However, because posed and natural expressions may differ from each other [16–18], it remains unclear whether the results of Bitti et al. [15] can be transferred to real interactive situations. Other studies have reported the production of facial expressions that show mental states related to thinking, such as uncertainty and concentration. For example, Krahmer and Swerts [19] found that when adult speakers were uncertain, they were more likely to produce a combination of facial movements, such as lip corner depression, lip stretching, or lip pressing, as well as eye widening and brow movements. Hübscher et al. [20] also revealed that adults who participated in a quiz game and were

uncertain were likely to wrinkle their nose, squint their eyelids, press/stretch their lips, and pull their lip corners down. From the perceivers' view, Rozin and Cohen [21] indicated that facial actions that signal thinking-concentration involve the eyes and eyebrows, specifically the narrowing of the eyes and lowering of the eyebrows. These findings suggest that the thinking face mainly includes eye and brow movements, with mouth movements occurring in some cases. However, it is unclear whether similar expressions are reproduced when participants are asked complex questions.

1.1.3 Uncanny Valley for Android Robots

An unnatural appearance or behavior in androids can lead to negative evaluations. This phenomenon is known as the uncanny valley [22, 23]. Uncanny valley effects are effectively measured using combinations of self-report scales of negative subjective experience (e.g., eerie and creepy) and ratings of human resemblance (humanlike, realistic, natural [22, 24]). More natural android behavior may reduce the uncanny valley effect by making the android appear less eerie and more human-like; for example, emotional expressions that lack detail appear uncannier [25]. Furthermore, a mechanical robot that exhibits incomplete (as opposed to no or full) nonverbal behavior in a social situation appears more uncanny [26]. Thus, a lack of expected social behavior may push an artificial entity into an uncanny valley. However, to date, no studies have investigated whether appropriate facial reactions in social situations mitigate the uncanny valley effect in androids. Uncanniness may be caused by the absence of certain facial expressions that are otherwise expected by social norms; in particular, since a thinking face is expected to follow a question, a lack thereof may appear uncannier.

1.2 The Present Set of Studies

This study aims to explore the production and perception of the thinking face when asked a complex question. Specifically, Study 1 tested how individuals express facial movements when thinking about something. We compared the facial movements produced under two conditions: the thinking condition, in which participants were asked a question and thought about their answer, and the non-thinking condition, in which participants mentally counted for 2 s. Furthermore, Study 2 explored whether facial patterns obtained in Study 1 could induce the perception that an android is thinking and that its behavior is appropriate to the context. For study 2, we implemented the thinking face pattern in the android robot Nikola that was validated to show human-like, rich facial actions [27]. It was expected that the thinking face obtained in the current study would be useful for engineering human–robot interaction, including its contribution to overcoming the uncanny valley. Study 3 further checked

the utility of the thinking face obtained in the current study by comparing a non-humanoid robot, which is a robot that does not have any human-like appearance.

For Study 1, in light of previous studies [1, 3, 12, 13, 15], we hypothesize that the thinking condition would show a more averted gaze as well as the raising of the eyebrows, lowering of the eyebrows, raising of the lower lip, and pulling of the lip corner compared to the non-thinking condition. However, because changes in several facial components and eye gaze could co-occur, we used non-negative matrix factorization (NMF) [28] to analyze these co-occurrences in an exploratory manner to determine if they produced expressions that corresponded to the thinking condition.

NMF, which helps obtain interpretable features in a low-dimensional space, has been applied to the time series data of facial movements [29–31]. Study 1, which included the hypotheses, exclusion criteria, and analysis plan, was preregistered (https://osf.io/zhkv5). In Study 2, it was hypothesized that the facial patterns obtained in Study 1, when implemented in an android, would result in attributions regarding it behaving more "being in thought," "human-likely," "appropriately," and less "creepy" than a neutral face (no facial movements).

2 Study 1

Study 1 was conducted to investigate the facial patterns that are obtained when an individual is "thinking about something." The current study compared facial movements produced under two conditions: the thinking condition, in which participants were asked a question and thought about their answer, and the non-thinking condition, in which participants counted for 2 s in their minds.

2.1 Methods

2.1.1 Participants

Participants were recruited by a temp agency under the following criteria: aged 20–30 years; whose native language was Japanese; who agreed to allow us to use the acquired facial data for research purposes; had no disease or psychological symptoms; and were in good physical condition and willing to wear masks and disinfect their hands to prevent COVID-19 infection. The experiment was performed in accordance with the COVID-19 Prevention Manual of RIKEN (https://osf.io/48suf). Forty adult Japanese individuals ($N = 40$) participated in this study (20 females and 20 males; mean age (SD) = 22.80 (3.36)). The sample size was determined based on a priori power analysis using G*Power 3.1.9.7 [32] with $d = 0.5$ (medium effect [33]; paired t-test: two-tailed). Written informed consent was obtained from each participant

before the start of the investigation, in line with the protocol approved by the Ethical Committee of RIKEN (Protocol number: W2022-067). They were paid 13,000 JPY for their time.

2.1.2 Procedure

In the current study, the experimenter randomly asked the participants the following six questions:

1. Without getting distracted, count for 2 s in your mind and say "yes" when you have finished (only count the numbers: control condition).
2. How has politics been affected by the expansion of force facilitated by technological innovation and dispersion? (Political question: thinking).
3. How will the pandemic change our lives in the future? (Prediction of life under the pandemic: thinking)
4. Multiply 9 and 12 (Calculation: thinking).
5. How would you deal with an android robot who looks completely human? (Question about an android: thinking)
6. What would you do if you got 100 million yen? (Individual question: thinking)

To control for the influence of prosody, the question was spoken using voice-reading software (VOICEPEAK). The original Japanese voices were uploaded to the OSF (https://osf.io/wr6ze).

The participants were tested individually and were instructed to respond as naturally as possible, considering the camera in front of them to be their conversational partner. Before the recording, the participants were asked to remove their masks, and if they were wearing glasses, they were also asked to remove them so that their facial patterns could be recorded clearly. The participants were seated, facing an RGB camera (RealSense D435, Intel: 30 fps). They were instructed to answer the six questions, and the camera was just to be saved for their answers for analysis.

In the debriefing sessions, participants were informed about the purpose of the current study to investigate the facial patterns behind the thinking face and were given the choice to either consent to having their facial expressions recorded or have us delete the data. After the debriefing, all participants gave their consent.

2.1.3 Statistical Analysis

The analysis targets were from the utterance of the question to the utterance of the answer, excluding verbal collateral signals (like "uh") from the recorded videos. For all facial videos (N = 240), we extracted frame-level facial action unit (AU) intensities on a 5-point scale using the automated facial

action detection system OpenFace (version 2.2.0) [34, 35]. Ekman et al. [36] developed the Facial Action Coding System (FACS) as an objective, comprehensive, and anatomy-based system to describe visible facial movements (e.g., Ekman and Rosenberg [37]). For example, moving the zygomatic major muscle to pull the corners of the lips is described as a "lip corner puller" (AU12). Based on the analysis by FACS, OpenFace can detect the following 17 AUs: 1 (inner brow raiser), 2 (outer brow raiser), 4 (brow lowerer), 5 (upper lid raiser), 6 (cheek raiser), 7 (lid tightener), 9 (nose wrinkler), 10 (upper lip raiser), 12 (lip corner puller), 14 (dimpler), 15 (lip corner depressor), 17 (chin raiser), 20 (lip stretcher), 23 (lip tightener), 25 (lips parts), 26 (jaw drop), and 45 (blink). OpenFace also provides eye direction using eye-gaze estimation [38]. The gaze direction was extracted by examining eye movements along the x- and y-axes, representing the up, down, left, and right directions. The estimated gaze and AU information obtained using OpenFace were regarded as the measured variables. Additionally, we applied the dimensional reduction approach using NMF because it would be redundant to use all the measured data (i.e., gaze and each AU) as individual dependent variables. The factorization rank was determined using cophenetic coefficients [39], dispersion coefficients [40], and RSS [41].

To compare the differences in the NMF factor scores between the thinking and non-thinking conditions, we performed paired t-tests. For each individual, mean values were calculated from all frame-level data. As a sub-analysis, to investigate whether the content of questions affected nonverbal information, NMF factor scores in the thinking condition were compared using six levels (types of questions) of repeated measures analysis of variance (ANOVA) and multiple comparisons using Shaffer's modified sequentially rejective Bonferroni procedure. We performed this analysis because it can be argued that it remains unclear whether each question type has specific facial patterns and whether the facial pattern in the control condition was really regarded as a non-thinking condition (more precisely, less thinking compared to other conditions). All analyses were performed using R statistical software (version 4.1.2; https://www.r-project.org/) and the "anovakun," "compute.es," "NMF," and "tidyverse" packages [42–45]. In keeping with open science practices that emphasize the transparency and replicability of results, all data and codes in the current study have been made available online (https://osf.io/wr6ze).

2.2 Results

The average time required to reach a response was calculated for six conditions. The results showed that the control condition and the calculation condition had the shortest duration (2.97 s and 3.01 s), followed by the individual question
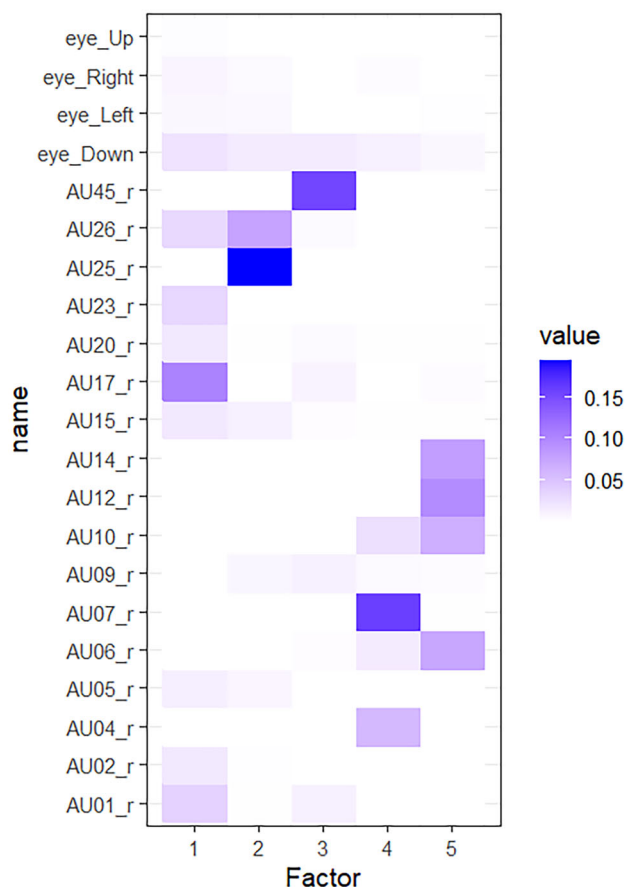

Fig. 1 Heatmap of each component's loadings for facial expressions of thinking. Note: Value colors represent each facial movement's contribution to factor scores

(3.67 s), the question about the android (4.47 s), the prediction of life during the pandemic (6.06 s), and finally the political question, which had the longest duration (12.28 s). The time for the control condition was shorter than for the other conditions. In addition, we recognized the possibility that the shorter duration of the control question might bias the facial reactions in a certain way. To confirm this possibility, we collected new data ($N = 14$) and compared the facial reactions in the "2-s counting condition" and "7-s counting condition." The results showed that there were no significant differences in the 17 facial movements observed between them ($ts < 2.05$, $ps > 0.06$). Therefore, we decided to consider the "simple counting number condition" as the control condition. The details of this preliminary study have been made available in the Supplementary Material.

We used NMF to identify facial patterns. The results revealed five facial patterns for all conditions (Fig. 1, Supplemental Fig. 1). Figure 1 shows the relative contribution of each AU to the independent components. Component 1 indicated raising of the chin (AU17), other slight facial movements (tightening of the lips, AU23; inner brow raiser,

AU1), and gazing down. The results of Component 2 indicated opening the mouth (AU25, AU26), whereas those of Component 3 suggested that blinking (AU45) was the main contributor. Component 4 was related to the narrowing of the eyes (AU7) and the lowering of the brows (AU4). The results of Component 5 correspond to smiling (AU12) along with raising of the cheeks (AU6). Furthermore, Component 5 included upper lip-raising (AU10) and dimpling (AU14); however, these AUs can be interpreted as confusion of AU12 in the automated action coding detection system [30, 46].

In the subsequent analysis, the average scores of the facial patterns for each condition and participant were treated as individual data points. Note that the target range of the frames may have varied.

To clarify the facial patterns obtained from individuals who are "thinking about something," we compared the difference in facial pattern intensities between thinking and non-thinking conditions using paired t-tests. Compared to the control condition (just counting numbers), all facial pattern intensities increased under the thinking condition (Table 1).

To investigate whether there were differences in the content of the conditions for each facial pattern, we conducted a repeated-measures ANOVA with the type of question (six levels) as a factor. For Component 1 (AU1 + 17 + 23 + gazing down), there was a statistically significant difference between questions ($F(5, 39) = 16.29$, $p < 0.001$). Multiple comparisons showed that only the counting condition was significantly lower than the other conditions (ts > 5.50, ps < 0.001) except for the calculation condition ($t = 2.80$, $p = 0.05$). Additionally, the intensity of Component 1 was significantly lower in the calculation task than in the political and COVID-19-related questions (ts > 3.57, ps < 0.007). Political questions exhibited a significantly higher intensity in Component 1 than personal questions ($t = 3.89$, $p = 0.004$). For Component 2 (AU25 + 26), there was a significant difference between questions ($F(5, 39) = 22.33$, $p < 0.001$). Multiple comparisons indicated that the counting condition was significantly lower than all other conditions ($t > 3.82$, ps < 0.001). The calculation condition was also significantly lower than the other remaining conditions (ts > 2.90, ps < 0.05). For blinking (Component 3), there was a significant difference between questions ($F(5, 39) = 3.93$, $p = 0.002$). Multiple comparisons indicated that the counting condition was significantly lower than the individual, political, and COVID-19-related questions (ts > 3.33, ps < 0.02). For Component 4 (AU4 + 7), there was a significant difference between the questions ($F(5, 39) = 6.88$, $p < 0.001$). Multiple comparisons indicated that COVID-19-related and political questions significantly contributed more to the intensity of Component 4 than the calculation and counting tasks (ts > 3.00, ps < 0.05). The COVID-19-related question was also significantly higher than that of the android question ($t = 3.25$, $p = 0.02$). Regarding smiling (Component 5), there was a significant

difference between the questions ($F(5, 39) = 4.76$, $p < 0.001$). Multiple comparisons indicated that the intensity of Component 5 in the counting condition was significantly lower than the other conditions (ts > 3.36, ps < 0.02) except for the calculation condition ($t = 2.43$, $p = 0.20$).

In summary, consistent with the results of the t-tests, most of the results showed that each facial pattern was expressed more strongly in the thinking condition than in the control condition.

To confirm the structure of facial patterns, we examined the correlation between each component at the frame level. Only correlations with effect sizes of 0.30 or higher, which can be considered a medium effect size [23], are reported below. We found a negative correlation between components 1 and 2 ($r = -0.30$), as well as a positive correlation between components 4 and 5 ($r = 0.30$).

2.3 Discussion

Study 1 aimed to clarify the facial patterns in human participants corresponding to "thinking about something." The results produced five facial patterns for thinking faces. More specifically, compared with the counting task, the situation in which participants had to think and answer questions induced raising of the chin/inner brows, tightening of the lips, and gazing down (Component 1). Additionally, thinking situations elicited opening of the mouth (Component 2), blinking (Component 3), furrowing (Component 4), and smiling (Component 5). These facial patterns are consistent with previous findings (e.g., Bavelas and Chovil [1]; Bitti et al. [15]; Chovil [3]; Goodwin and Goodwin [12]; Heller [13]).

Opening the mouth (Component 2) is a novel form that has not been previously reported in the relevant literature. From a data-driven perspective based on effect sizes, this movement can be considered the most significant component, but it has been noted that this movement might heighten the perception that it is deliberate [30]. There were concerns regarding the attributions of intentionality and arbitrariness when implementing this movement in androids. Although Component 1 (raising the chin/inner brows, tightening the lips, and gazing down) also exhibited a large effect size, its loadings spanned multiple facial actions, making it difficult to interpret it as a fixed facial pattern. AU17 (raising the chin) and 23 (tightening the lips) versus AU26 (jaw drop) represent conflicting actions. In addition, AU17 computed by OpenFace may indicate noise when AU25 and 26 (opening the mouth) were present, according to previous findings by Namba et al. [30]. Furthermore, our correlation results suggest that Component 1 may be noise that covaries with Component 2.

Importantly, the furrowing of the brow and narrowing of the eyes observed in Component 4 align with previous

Table 1 Results of paired t-tests for the difference in facial patterns between thinking and control conditions

Component	Mean$_{thinking}$	Mean$_{control}$	t-value	p-value	Hedges'g
1	2.82	1.45	10.64	< .001	2.36
2	2.52	0.83	12.45	< .001	2.76
3	2.08	1.56	3.81	< .001	0.84
4	2.22	1.15	4.55	< .001	1.01
5	2.99	0.94	4.61	< .001	1.02

research findings. In particular, Component 4 was in accordance with the facial patterns observed in both production (e.g., Bitti et al. [15]) and perception [21] studies. This indicates the robustness and replicability of the furrowed thinking face. In exploring facial patterns made by an android to induce the perception of "being in thought," it is more useful to adopt facial components established in previous research to be recognized by humans as facial actions that signal thinking-concentration.

Therefore, we selected Component 4 as the main thinking face pattern for the human–robot interaction study. In Study 2, we aimed to clarify whether implementing the observed thinking faces in an android can facilitate natural human–robot interaction. More specifically, we confirmed whether the facial patterns related to furrowing implemented in the android result in attributions of it behaving "being in thought" and "appropriately." In addition, as a thinking face becomes more complex, the perception of "being in thought" may be amplified due to the increase in facial information. Therefore, it is also necessary for the natural human–robot interaction to investigate whether the combination of Components 4 and 5 (furrowed smiling), for which a positive correlation was observed (it can be determined that they can co-occur), conveys a state of thinking.

3 Study 2

Building on the results of Study 1, Study 2 was conducted to examine whether the thinking face can induce the perception of "being in thought" and "appropriateness" compared to a neutral face (no facial movements) when applied to an android robot. In Study 2, two types of facial patterns were adopted: narrowing of the eyes and furrowing of the brow as a thinking condition, and adding a smile to that facial pattern. It was expected that, compared to the neutral face, these facial patterns would evoke perceptions of "being in thought" and "appropriateness."

Furthermore, as Sect. 1.1.3 suggested, it was expected that an android depicting a thinking face would be less eerie and more human-like than an android with a neutral facial expression.

Recent studies have indicated that contextual information is important for the interpretation of facial expressions [47–49]. Considering the connection to Study 1, the context in which the robot was asked the same questions as in Study 1 was established.

3.1 Methods

3.1.1 Participants

A total of 89 crowdsourced workers (37 women and 47 men; age range = 21–64 years, mean = 41.02, SD = 7.69) agreed to participate in a survey via Crowdworks (CW: www.crowdworks.jp), and all participants were Japanese. The validation of CW participants has already been confirmed by Majima et al. [50] and is aligned with that of normal participants in behavioral experiments. Informed consent on the CW platform was obtained from each participant before the investigation, in line with the protocol approved by the Ethical Committee of RIKEN (Protocol number: W2022-067). This study was conducted in accordance with the ethical guidelines of our institute and the Declaration of Helsinki. After completing the experimental task, the participants received 440 JPY for completing a 10-min survey.

3.1.2 Stimuli

The android robot Nikola was used because of its ability to simulate realistic facial patterns [27]. Nikola has 29 pneumatic actuators on its face to validly reproduce 17 AUs, including all the AUs of interest in this study, as well as six electrical actuators for head and eyeball control. Nikola was programmed to show different expressions, and three frontal videos of Nikola's face were captured for each expression. Six video clips were created depicting the three facial patterns (the furrowed face, furrowed face with a smile, and neutral face; Fig. 2) as responses to the two types of questions. To address the issue of the forehead appearing slightly larger and unnatural because of the lack of hair and large number of actuators, the android's head was covered with a hat during the recording to create a more natural appearance. The content of the questions included two that induced the thinking face in Study 1 (the political question and prediction
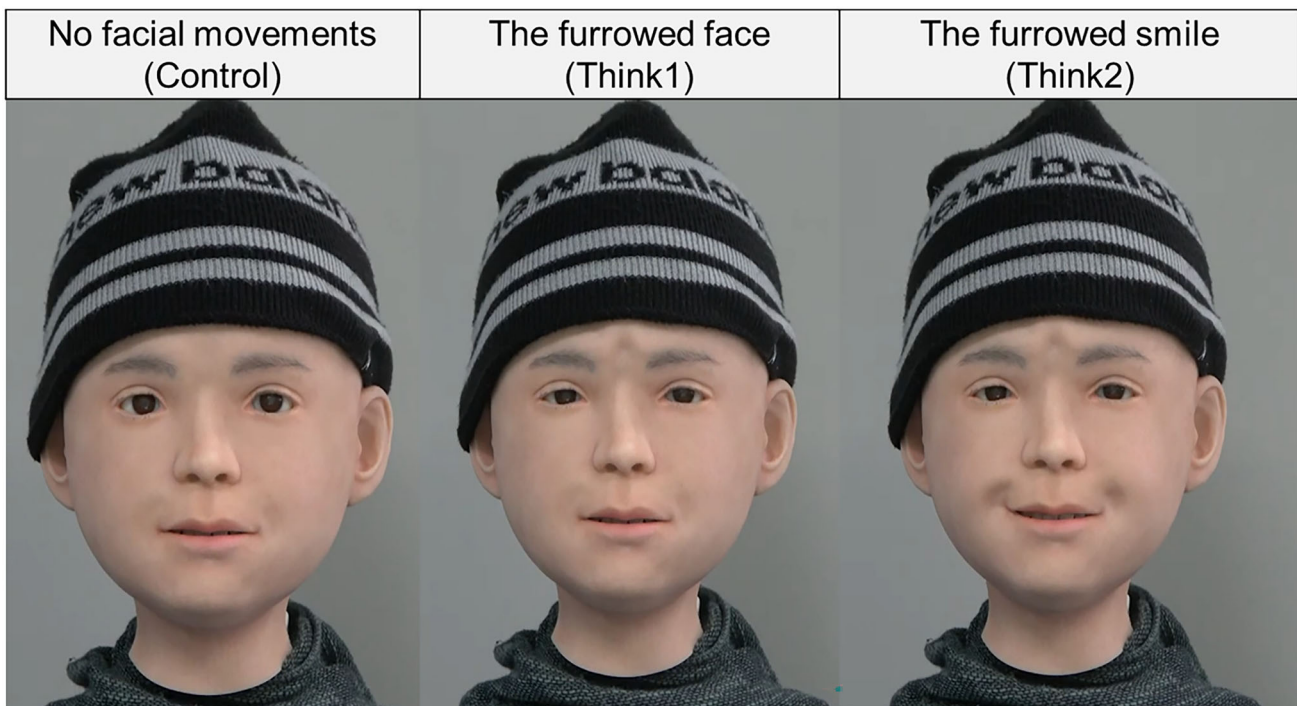
No facial movements (Control)	The furrowed face (Think1)	The furrowed smile (Think2)

Fig. 2 Illustrations of the facial expressions of three conditions produced by the android Nikola

of life during the pandemic). The activated AUs included 4 (brow lowerer) and 7 (lid tightener) for the furrowed face; 4, 6 (cheek raiser), 7, and 12 (lip corner puller) for the furrowed smiling face; and nothing for the neutral face. The videos were recorded using a digital web camera (HD 1080P; Logicool, Tokyo, Japan). All videos are available at OSF (https://osf.io/wr6ze).

3.1.3 Procedure

The experiment was conducted using the Qualtrics online platform (Seattle, WA, United States). The video clips of Nikola's facial patterns were presented on the monitor individually in a random order, and rating tasks for "genuineness" [51], "eeriness/human-likeness" [52, 53], "thinking about the answer," and "appropriateness" [54] were presented below each clip. The total number of videos viewed and rated by all participants was six, which included within design, which included a within-participant factor (3: control, furrowed thinking, and smiled thinking). Participants were also asked, "What psychological states does this robot seem to have?" They could answer using a free description. No time limits were set, and no feedback on the response was provided. An image from each clip was presented randomly. "Genuineness" was measured by asking participants, "How genuine does this facial reaction look to you?" using a Likert scale ranging from "completely fake (−3)" to "completely genuine (+ 3)". "Eeriness" and "human-likeness"

were measured using 2 and 3 items, respectively, as "This robot is creepy/eerie" and "This robot is unnatural/human-like/machine-like," ranging from 1 "not at all" to 7 "very much so." Eeriness and human-likeness ratings were selected based on previous studies and recommendations [22, 24]. The "thinking about the answer" component was measured by the following item: "This robot is thinking about the answer," ranging from 1 "not at all" to 7 "very much so." Six "appropriateness" items were measured using the same 7-point Likert scale. The contents for "appropriateness" are as following: "this expression is appropriate," "this display is odd," and "this expression fits the situation." The current study applied a 7-point Likert scale because a positive relationship exists between the number of scale points and the reliability of the measurement [55], and seven is a reasonable number of categories [56]. The use of a scale with more than seven points may be less meaningful to raters [57]. In total, each participant made 14 evaluations for one clip.

3.1.4 Statistical Analysis

To clarify the relationship between facial patterns and rating scores, a hierarchical linear model was used to control for differences between each participant and stimuli. The models for each rating score were as follows:

$$\text{Rating scores} = B_{intercept} + B_{Thinking1} + B_{Thinking2} \qquad (1)$$

$$B = \gamma_{\mathrm{intercept}} + \gamma_{stimuli} + \gamma_{participants} \qquad (2)$$

All analyses were performed using R statistical software (version 4.1.2; https://www.r-project.org/) and the "lmerTest," "psych," and "tidyverse" packages [45, 58, 59]. In keeping with open science practices that emphasize the transparency and replicability of results, all data and codes in the current study have been made available online (https://osf.io/wr6ze).

For free description data, the authors and one part-time researcher classified the data into 14 categories as nominal variables. The categories were: "1: response to a difficult question," "2: unconcerned," "3: cannot understand," "4: depressed feeling," "5: thinking seriously," "6: positive feeling," "7: aggressive feeling," "8: scorn," "9: confidence," "10: brewing mischief," "11: eeriness, unnatural," "12: bitter smile," "13: inattention," and "14: others (just the description of the facial movement)." These 14 categories were discussed and finalized by the three authors (SN, SN, and WS) while coding the data. This discussion was prompted by the realization that the reports were more complex and diverse than had been anticipated. The mean agreement regarding labeling (Cohen's $\kappa = 0.87$) was sufficiently high to suggest intercoder reliability.

3.2 Results

3.2.1 Perception of Thinking

The hierarchical linear model revealed that the thinking faces significantly heightened the score of "thinking about the answer" (Furrowed face: B= 2.14, $t = 7.72$, $p = 0.006$; Furrowed face with smile: B = 1.01, $t = 3.39$, $p = 0.03$; Fig. 3). This indicates that both thinking faces were perceived as engaged in thought.

3.2.2 Perception of Genuineness

The hierarchical linear model showed that the furrowed face significantly increased the score of "genuineness" (B = 1.28, $t = 8.58$, $p < 0.001$), but the furrowed smiling face did not show a significant difference compared to the neutral face (B = 0.39, $t = 1.91$, $p = 0.06$).

3.2.3 Perception of Eeriness

For "eeriness" ($\alpha = 0.90$), the results indicated that the furrowed face significantly lowered the score of "eeriness" (B = -0.36, $t = 2.57$, $p = 0.012$), but the furrowed smiling face significantly heightened it compared to the neutral face (B = 0.81, $t = 4.04$, $p = 0.009$).

3.2.4 Perception of Human-likeness

For "human-likeness" ($\alpha = 0.83$), the results revealed that the furrowed face significantly increased the score of "human-likeness" (B = 1.11, $t = 7.22$, $p = 0.002$), but the furrowed smiling face did not make a significant difference compared with the neutral face (B = -0.35, $t = 2.07$, $p = 0.085$).

3.2.5 Perception of Appropriateness

For "appropriateness" ($\alpha = 0.96$), the results revealed that the furrowed face significantly increased the score of "appropriateness" (B = 1.18, $t = 5.41$, $p = 0.003$), but the furrowed smiling face did not make a significant difference compared to the neutral face (B = -0.23, $t = 1.15$, $p = 0.26$).

In sum, the results revealed that the furrowed face can induce perceptions of "thinking about the answer," "genuineness," "low-eeriness," "human-likeness," and "appropriateness" compared to a neutral face, but the furrowed smiling face did not do so.

3.2.6 Free Description Data

Table 2 lists the response categories for each condition. The results in Table 2 suggest that the neutral face gave participants the impression of being "unconcerned." The furrowed face expressed "thinking seriously." However, this facial pattern also caused negative feelings such as "depressed feeling" and "aggressive feeling." At the same time, the furrowed smiling face expressed a "positive feeling" along with "scorn" and "brewing mischief."

Therefore, the free description results found that the neutral face was perceived as being "unconcerned" and the furrowed face as "thinking." At the same time, the furrowed face evoked negative emotions, whereas the furrowed smiling face evoked positive or condescending emotions.

3.3 Discussion

Study 2 aimed to investigate whether the "thinking face" expression can induce attribution among viewers that an android robot was "thinking" and that its behavior was "appropriate" rather than a neutral facial expression (no facial movements). The results revealed that the furrowed face elicited the attributions of "thinking about the answer," "genuineness," "low eeriness," "human-likeness," and "appropriateness" compared to the neutral face, while the furrowed smiling face did not do so. Additionally, the free description results showed that the furrowed face induced negative feelings and was perceived as "thinking seriously," and the furrowed smiling face evoked mainly positive or condescending emotions.

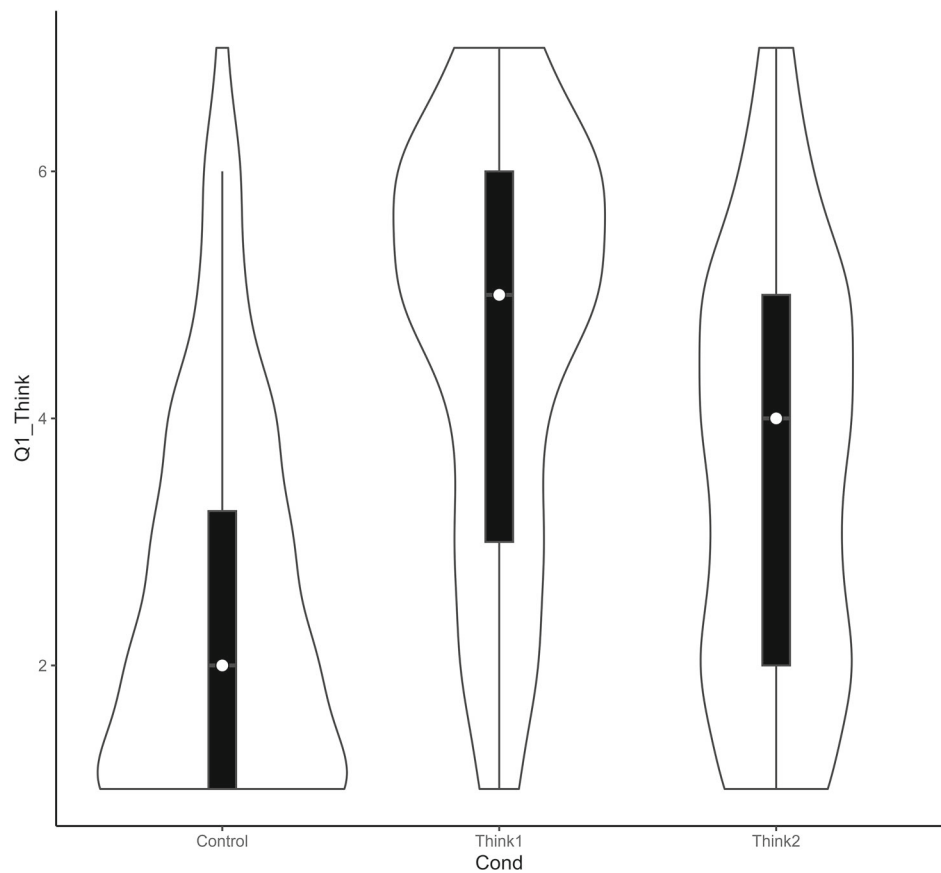**Fig. 3** Violin plots of the "thinking about the answer" variables

Table 2 Frequencies of each label from free description data

Label category													
Condition	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Control	25	79	10	6	16	3	6	0	1	0	2	0	2	18
Think1	25	4	3	43	34	2	41	0	0	0	0	0	3	13
Think2	18	8	1	5	4	33	19	20	13	26	0	5	2	14

"1: response to difficult question," "2: unconcerned," "3: cannot understand," "4: depressed feeling," "5: thinking seriously," "6: positive feeling," "7: aggressive feeling," "8: scorn," "9: confidence," "10: brew mischief," "11: eeriness, unnatural," "12: bitter smile," "13: inattention" and "14: others (just the description of the facial movement)"

Although Study 2 only included two facial patterns as the thinking face, it may be evident that all facial displays that humans are expected to express in actual question-and-answer situations do not necessarily induce the attributions of "genuineness," "low eeriness," and "appropriateness." Studies on facial expressions have repeatedly indicated that production and perception issues are distinct entities [60, 61]. In that sense, it is noteworthy that the furrowed face as one of the thinking faces elicited valid attributions ("thinking," "genuineness," "low eeriness," "human-likeness," and "appropriateness") for the natural human–robot interaction.

In the free-description data, several participants recognized the mental state and emotions of the robot through the portrayal of the thinking face. Certain facial expressions have the potential to convey multiple types of information, including representations of what the world is like [2, 62]. Given that the emotional information in facial expressions tends to be prioritized over other types of information [62, 63], it is reasonable to expect such results.

Regarding the furrowed smiling face, positive emotions were attributed to it in the free-description data. However, while the furrowed face decreased the eeriness ratings, the furrowed smiling face increased the eeriness compared to

the no-facial movement condition. One possible interpretation is that there might be a sense of aversion that humans associate with a robot creating a positive expression, such as a smile. Given that the furrowed smile did not increase genuineness, appropriateness, or human likeness compared with the neutral face, the furrowed smiling face may not have been recognized as a proper thinking face. Alternatively, the smile may have been interpreted as an expression of a positive emotion and thus deemed an inappropriate response to a complex question. In any case, an inappropriate facial reaction would thus be considered eerier than no reaction at all, potentially caused by a mismatch between the social situation and the facial reaction. Another consideration is related to hardware issues, as the intensity of the lip corner puller in the Nikola android may have been too small [27], resulting in a physically unnatural state. Further investigation is required to gain deeper insight into these findings.

Taken together, Study 2 showed that the thinking face (furrowed face) elicited perceptions related to "being in thought," "genuineness," "low eeriness," and "appropriateness." In Study 3, the current study would further examine the complemental applicability of the thinking face obtained in this study by comparing it with expressions that appear to be thinking by non-humanoid robots.

4 Study 3

In Study 1, we identified the facial patterns expressed by individuals engaged in answering complex questions (i.e., thinking face) and clarified in Study 2 how implementing these observed thinking faces in an android can produce greater perceptions related to natural human–robot interactions. The primary objective of Study 3 was to further prove these perceptions in question–answer situations.

The secondary objective was to compare the effect of showing "being in thought" between an android and a chatbot. Artificial agents have many indicators that work in line with the thinking face to facilitate human–robot interaction [64–67]. For example, an increasing sequence of dots is a common indicator used to indicate that a robot is loading a response [66]. Since the rise of ChatGPT (OpenAI, 2023), chat-based human–robot interactions have increased in number and become more routine than before. Therefore, comparing human responses to thinking faces in an android and chat-based loading representations could provide additional insight into the findings of the current study.

Accordingly, we conducted Study 3 to exploratively investigate humans' perceptions of thinking vs. control faces in an android and thinking (i.e., an increasing sequence of dots) vs. control displays of a chatbot in question–answer situations.

4.1 Methods

4.1.1 Participants

A total of 40 Japanese university students (32 women and 8 men; age range = 19–28 years, mean = 20.92, SD = 2.14) agreed to participate in a survey. Informed consent was obtained from each participant before the investigation, in line with the protocol approved by the Ethical Committee of RIKEN (Protocol number: W2022-067). This study was conducted in accordance with the ethical guidelines of our institute and the Declaration of Helsinki. After completing the experimental task, the participants received Amazon gift cards (300 JPY) for completing the survey.

4.1.2 Stimuli

Study 3 also used the android robot Nikola as the test android. Nikola was programmed to show two facial patterns: the combination of AU4 (brow lowered) and AU7 (lid tightener) for the thinking face and no facial movements for the neutral face. We set up the hypothetical question–answer situation in which a human asked two types of questions ("Please tell me about a delicious sushi restaurant in Tokyo." or "Please tell me how the development of technology has changed Japanese politics."), and after a 2.5-s (thinking/loading time), the robots answered, "Of course." During this 2.5-s waiting period, one condition showed a thinking face in Nikola, and the other condition showed a neutral face. The videos were recorded using a digital web camera (HD 1080P; Logicool, Tokyo, Japan). To control for the influence of prosody, all sounds were spoken using voice-reading software (VOICE-PEAK).

For the non-humanoid robot condition, we created a chat-based interaction. As the human typed, a question voice (same as the android stimuli) was played, and after waiting for the same 2.5 s, the robot responded, "Of course." During the 2.5-s waiting period, two types of conditions were prepared: a condition in which nothing was presented and a condition in which dots were continuously presented (Fig. 4). Eight video clips were created depicting the two agents (Nikola and the chatbot) with the two thinking patterns (thinking and a neutral face) as responses to the two types of questions. All videos are available at OSF (https://osf.io/wr6ze).

4.1.3 Procedure

The experiment was conducted using jsPsych [68]. Data were saved in Cognition (https://www.cognition.run/), which is a platform with which experiments can be hosted. The video clips of Nikola's facial patterns and chat-based interactions were presented on the monitor individually in a random

Fig. 4 Illustrations of the two conditions produced for the chat-based human–robot interaction
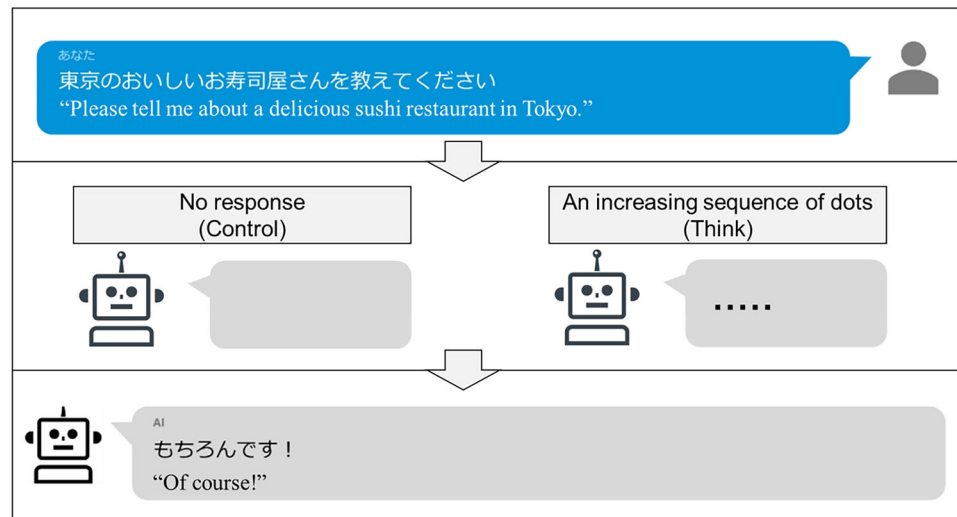
order. Similar to Study 2, rating tasks for "eeriness/human-likeness," "thinking about the answer," and "appropriateness" were presented below each clip. The total number of videos viewed and rated by all participants was eight, which included within design, which included two within-participant factors (2 agents: Nikola and chatbot; 2 conditions: control and thinking face). No time limits were set, and no feedback on the response was provided. Each participant made 12 evaluations per clip.

4.1.4 Statistical Analysis

For an exploratory investigation of how each agent's thinking face is perceived, we conducted a 2×2 repeated measures ANOVA, with the factors agent (Nikola and chatbot) and condition (control and thinking) as within-subject factors. All analyses were performed using R statistical software (version 4.1.2; https://www.r-project.org/) with the "anovakun" and "tidyverse" packages [44, 45]. In keeping with open science practices, which emphasize the transparency and replicability of results, all data and codes in the current study have been made available online (https://osf.io/wr6ze).

4.2 Results

4.2.1 Perception of Thinking

For perceptions of "being in thought," the results indicated a significant main effect of the agents, $F(1, 39) = 27.54$, $p < 0.001$, partial $\eta2 = 0.41$. A significant main effect of the conditions was observed, $F(1, 39) = 5.92$, $p = 0.02$, partial $\eta2 = 0.13$. Meanwhile, the interaction between the agents and the conditions showed a trend towards significance, $F(1, 39) = 3.61$, $p = 0.065$, partial $\eta2 = 0.09$. A closer look at the interaction effect revealed that there was no difference
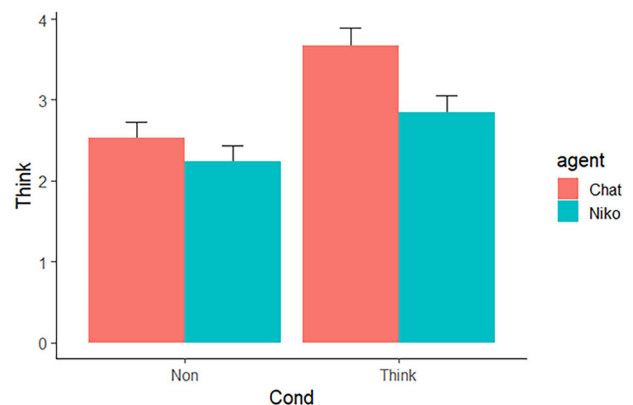

Fig. 5 Bar plot of the "thinking about the answer" variables

between agents in the neutral condition, $F(1,39) = 1.76$, $p = 0.19$, partial $\eta2 = 0.04$. While there was a significant difference between agents in the think condition, $F(1,39) = 7.26$, $p = 0.01$, partial $\eta2 = 0.16$.

In summary, regardless of the difference in agent type, the perception of "being in thought" increased in the thinking conditions (thinking face or increasing dots) compared to the control condition (no responses from the robots), but it became clear that the degree of increase was greater for the chatbot than for Nikola (Fig. 5).

4.2.2 Perception of Eeriness

For "eeriness" ($\alpha = 0.96$), the results indicated only one significant main effect of the agents, $F(1, 39) = 136.99$, $p = 0.000$, partial $\eta2 = 0.78$. Simply put, Nikola was judged to be eerier than the chat-based robot.

4.2.3 Perception of Human-likeness

For "human-likeness," reliability was too low ($\alpha = 0.37$). Thus, we used the human-likeness that excluded the item "this robot is unnatural" ($\alpha = 0.80$). The results indicated a significant main effect of the agents, $F(1, 39) = 12.82$, $p < 0.001$, $partial\ \eta2 = 0.25$. A significant main effect of the conditions was also observed, $F(1, 39) = 18.38$, $p < 0.001$, $partial\ \eta2 = 0.32$. While the interaction between the agents and the conditions showed a trend towards significance, $F(1, 39) = 3.13$, $p = 0.085$, $partial\ \eta2 = 0.07$. A closer look at the interaction effect revealed that there was a slight difference between conditions in the chatbot, $F(1,39) = 3.65$, $p = 0.06$, $partial\ \eta2 = 0.09$. While there was a significant difference between conditions in the android robot, $F(1,39) = 13.58$, $p < 0.001$, $partial\ \eta2 = 0.26$.

4.2.4 Perception of Appropriateness

For "appropriateness" ($\alpha = 0.88$), the results indicated only one significant main effect of the agents, $F(1, 39) = 136.99$, $p = 0.000$, $partial\ \eta2 = 0.78$. The chatbot was judged to be more appropriate than the android.

In sum, the results revealed that the thinking face (furrowed face in the android and increasing dots in the chatbot) can induce perceptions of "being in thought," and "human-likeness." However, the effect of thinking face was greater for chatbots in the case of "being in thought" and greater for androids in the case of "human-likeness."

4.3 Discussion

In Study 3, we compared human responses to thinking faces (i.e., furrowed face) in an android and to chat-based loading representation (i.e., an increasing sequence of dots). As a result, we found that the thinking face (furrowed face in android and increasing dots in the chatbot) can induce perceptions of "being in thought" and "human-likeness." However, the effect of thinking face was greater for chatbots in the case of "being in thought" and greater for androids in the case of "human-likeness."

The reason the perception of "being in thought" was greater for chatbots may be familiarity. As mentioned in the introduction to Study 3, indicators such as increasing dots that indicate loading information for users have become more frequent. Therefore, even in response to a simple expression such as "the number of dots increases," it is easy for participants to predict from experience that "this is the time to wait." How the observed result will change as androids become more common for people is an interesting avenue for future research.

The use of a thinking face on Nikola resulted in the android being perceived as more human-like. This new finding is consistent with Study 2. It can be said that human-like behavior (Study 1) has shown the ability to transform androids into beings that are more than machines. An android acting in accordance with expected and appropriate human social behavior is perceived as being more humanlike. The ability to react with appropriate social behavior may lead to the attribution of a human mind, which in turn increases anthropomorphizing.

For "Eeriness" and "Appropriateness," no significant effect was found for Nikola's thinking face compared to the neutral face, unlike in Study 2. This may have been due to the application of a within-participant design that included a chat-based interaction. It is easy to imagine that when participants compare familiar chat-based interactions to android interactions (which they are probably seeing for the first time), the latter will be perceived as strange regardless of the android's facial expression. Additionally, the difference in appropriateness may be explained by the fact that the chatbot was designed for text-based interaction, while Nikola seemed to be designed for conversational interaction. The appropriate period of silence in a conversation is said to be approximately 0.4 s on average [69]. In this study, the silence was 2.5 s long, more than six times as long. This extended silence may have been judged to be "inappropriate." Therefore, future research will need to deal with conversational expressions by non-humanoid robots that are more appropriate for comparison with Nikola.

Regarding Study 3, it should be noted that there is a spectrum of robots between Nikola and the chatbot. For example, the robot called EDDIE is both human- and animal-like [70], while Fukuda et al. [71] developed the robot called KAPPA, which looks like a Japanese character. There have also been android robots with many different properties compared to Nikola (e.g., ERICA: Glas et al. [72]; flobi: Hegel et al. [73]; IURO: Mirnig et al. [74]). Accumulating knowledge on human–robot interaction according to different purposes is crucial for the future, and this study could serve as a touchstone for this purpose.

5 General Discussion

In this study, we aimed to (a) identify the facial patterns expressed by individuals engaged in answering complex questions (i.e., thinking face: Study 1) and (b) clarify whether implementing these observed thinking faces in an android can facilitate natural human–robot interaction (Study 2 and Study 3). Study 1 identified five facial patterns of a person answering questions as thinking faces. Focusing on the furrowed face/furrowed smiling face among these facial patterns, Study 2 provided evidence that the furrowed face (without a smile) elicited perceptions related to "being in thought," "genuineness," "low eeriness," and "appropriate-

ness." Finally, Study 3 found that a furrowed face on an android enhances the perception that it is "more human-like," compared to the increasing dots in the chatbot.

The current study identified facial patterns that emerged when individuals contemplated their responses to questions. Subsequently, a subset of these expressions was implemented in an android, resulting in the attribution of "thinking about the answer." These findings provide insights into the natural interactions between humans and android robots. In particular, a situation in which a person asks a question to an android robot can become a common scenario in the future. If the robot expresses the facial pattern obtained in the current study, it can be considered the initial step toward natural human–robot interaction that humans perceive as human-like and appropriate, while reducing the uncanny valley effect through more natural social reactions.

This study is the first to show that a lack of (or inappropriate) facial reactions in a social context can increase eeriness and decrease human-likeness ratings, leading to an uncanny valley. These results are consistent with previous findings that a robot's incomplete social behavior can cause discomfort [26]. The results suggest that the uncanny valley effect can be extended to social situations, while the effect itself is typically considered to be caused by deviating or mismatching information in an entity's appearance [22]. This idea can be extended to a lack or mismatch between the social context and the reaction, as shown in this study. Thus, eeriness may be caused by domain-independent deviations or mismatches across multiple levels of perception, ranging from expectations of physical appearance to socially appropriate behaviors.

The current study investigated both the production and perception aspects of thinking faces, but their pragmatic functions were not fully elucidated. Hömke et al. [75, 76] indicated that listener blinks are perceived as communicative signals that directly influence speakers' communicative behavior in face-to-face communication. Thus, blinking, which had been observed as Component 3 and as part of the facial pattern related to thinking faces, may have the same pragmatic function. Future studies can garner great interest in attempting to understand the information that thinking faces, obtained in the current study, signals in human–robot interaction, and how it can potentially and pragmatically influence human behavior.

In Study 1, some facial patterns observed in previous studies, such as pulling the lip corners down [19, 20] and raising the eyebrows [14], were not observed. While there are contextual differences between studies, a possible factor that should be considered is cultural differences. In certain nonverbal communications, cultural differences can occur in both expression and perception [77]. Recent studies have accumulated evidence regarding cultural differences in production (e.g., Cordaro et al. [78]; Fang et al. [79])

and perception (e.g., Jack et al. [80]; Masuda et al. [81]; Namba et al. [82]). By further considering cultural aspects and expanding our understanding, the international applicability of the findings obtained in the current study can be more comprehensively appreciated, emphasizing their relevance for applying insights to android robots.

It is worth mentioning the social value provided by the present sets of studies. The current works revealed the relatively appropriate facial signals that Nikola can convey "being in thought" and "more human-like." However, Studies 2 and 3 used video clip stimuli exhibited by Nikola to participants, thus it was not socialized with participants. To achieve truly appropriate human–robot interaction in a social setting, in addition to realizing linguistic interactions (which are becoming feasible with the advent of generative AI), it might be necessary to establish mechanisms for sharing attention [83], along with the pragmatic facial movements examined in the current studies. For instance, people exhibit similar physiological responses when making eye contact with a robot as they do when making eye contact with a person [84]. This implies that a sense of shared experience might be crucial in social situations that foster more natural interactions. In addition to the expressive aspects explored in the current studies, the development of systems capable of sharing attention will be required in the future.

It should be noted that we could not provide insights into the temporal/dynamic information. The importance of temporal information has been repeatedly emphasized in the expression and interpretation of facial signals (e.g., Krumhuber et al. [85]; Krumhuber et al. [86]; Sato et al. [87]). However, this aspect could not be fully explored in this study as its focus was limited to spatial information. With further data extension and refinement of analytical approaches, future studies should examine temporal information. In addition, the existing automated assessment system for AUs enables the determination of AU intensity on a frame-by-frame basis. The use of an automated AU detection system has certain advantages. However, despite advancements in machine learning and artificial intelligence techniques in the field of affective computing [88], this system has flaws [89]. Therefore, replication studies using more sophisticated facial movement detection systems would be useful.

In conclusion, Study 1 involved the analysis of the facial movements of 40 participants as they responded to challenging questions, leading to the identification of five facial patterns associated with thinking faces. Building on Study 1, Study 2 specifically focused on the furrowing of the brows and narrowing of the eyes among the observed thinking facial patterns, which were then implemented in an android. The findings indicated that thinking faces enhanced perceptions of thinking, genuineness, human-likeness, and appropriateness, while reducing the uncanny valley effect. Moreover, analysis of the free description data revealed an

attribution of negative emotions to the thinking face. Study 3 also showed that just the trick of "expressing furrowed face in android during a part of a conversation" increases attributions of human-likeness when compared to text-based human–robot interactions, and further human-like behavior will enhance the knowledge obtained in the current study. This research lays the groundwork for achieving natural human–robot interaction.

Data availability The datasets presented in this study can be found in online repositories. The names of the repository/repositories can be found at: https://osf.io/wr6ze

Declarations

Conflict of interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bavelas J, Chovil N (2018) Some pragmatic functions of conversational facial gestures. Gesture 17:98–127. https://doi.org/10.1075/gest.00012.bav
2. Scarantino A (2017) How to do things with emotional expressions: the theory of affective pragmatics. Psychol Inq 28:165–185. https://doi.org/10.1080/1047840X.2017.1328951
3. Chovil N (1991) Discourse-oriented facial displays in conversation. Res Lang Soc Interact 25:163–194. https://doi.org/10.1080/08351819109389361
4. Sun J, Harris K, Vazire S (2020) Is well-being associated with the quantity and quality of social interactions? J Pers Soc Psychol 119:1478–1496. https://doi.org/10.1037/pspp0000272
5. Watson D, Clark LA, McIntyre CW, Hamaker S (1992) Affect, personality, and social activity. J Pers Soc Psychol 63:1011–1025
6. Ishi C, Minato T, Ishiguro H (2017) Motion analysis in vocalized surprise expressions. Interspeech. https://doi.org/10.21437/Interspeech.2017-631
7. Fukuda T, Jung MJ, Nakashima M et al (2004) Facial expressive robotic head system for human-robot communication and

8. its application in home environment. Proceedings of the IEEE 92:1851–1865. https://doi.org/10.1109/jproc.2004.835355
8. Kobayashi H, Hara F (1997) Facial interaction between animated 3D face robot and human beings. 1997 IEEE International conference on systems, man, and cybernetics computational cybernetics and simulation. https://doi.org/10.1109/icsmc.1997.633250
9. Lazzeri N, Mazzei D, Greco A, Rotesi A, Lanatà A, De Rossi DE (2015) Can a humanoid face be expressive? a psychophysiological investigation. Front Bioeng Biotechnol 3:64
10. Ekman P (1993) Facial expression and emotion. Am Psychol 48:384392
11. Stock-Homburg R (2022) Survey of emotions in human–robot interactions: perspectives from robotic psychology on 20 years of research. Int J Soc Robot 14(2):389–411
12. Goodwin MH, Goodwin C (1986) Gesture and coparticipation in the activity of searching for a word. Semiotica 62:51–75
13. Heller V (2021) Embodied displays of "doing thinking" epistemic and interactive functions of thinking displays in children's argumentative activities. Front Psychol 12:636671
14. Nota N, Trujillo JP, Holler J (2021) Facial signals and social actions in multimodal face-to-face interaction. Brain Sci 11:1017. https://doi.org/10.3390/brainsci11081017
15. Bitti PER, Bonfiglioli L, Melani P, Caterina R, Garotti P (2014) Expression and communication of doubt/uncertainty through facial expression. Ricerche di pedagogia e didattica. J Theor Res Educ 9:159–177
16. Hess U, Kleck RE (1990) Differentiating emotion elicited and deliberate emotional facial expressions. Eur J Soc Psychol 20:369–385. https://doi.org/10.1002/ejsp.2420200502
17. Schmidt KL, Bhattacharya S, Denlinger R (2009) Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. J Nonverbal Behav 33:35–45. https://doi.org/10.1007/s10919-008-0058-6
18. Namba S, Makihara S, Kabir RS, Miyatani M, Nakao T (2017) Spontaneous facial expressions are different from posed facial expressions: morphological properties and dynamic sequences. Curr Psychol 36:593–605. https://doi.org/10.1007/s12144-016-9448-9
19. Krahmer E, Swerts M (2005) How children and adults produce and perceive uncertainty in audiovisual speech. Lang Speech 48:29–53. https://doi.org/10.1177/00238309050480010201
20. Hübscher I, Vincze L, Prieto P (2019) Children's signaling of their uncertain knowledge state: prosody, face, and body cues come first. Lang Learn Dev 15:366–389. https://doi.org/10.1080/15475441.2019.1645669
21. Rozin P, Cohen AB (2003) High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of americans. Emotion 3:68–75. https://doi.org/10.1037/1528-3542.3.1.68
22. Diel A, Weigelt S, MacDorman KF (2021) A meta-analysis of the uncanny valley's independent and dependent variables. J Hum Robot Interact 11:1–33. https://doi.org/10.1145/3470742
23. Mori M, MacDorman K, Kageki N (2012) The uncanny valley [from the field]. IEEE Robot Automat Mag 19:98–100. https://doi.org/10.1109/MRA.2012.2192811
24. Ho C-C, MacDorman KF (2017) Measuring the uncanny valley effect: refinements to indices for perceived humanness, attractiveness, and eeriness. Int J Soc Robot 9:129–139. https://doi.org/10.1007/s12369-016-0380-9
25. Tinwell A, Grimshaw M, Nabi DA, Williams A (2011) Facial expression of emotion and perception of the uncanny valley in virtual characters. Comput Hum Behav 27:741–749. https://doi.org/10.1016/j.chb.2010.10.018
26. Thepsoonthorn C, Ogawa K-i, Miyake Y (2021) The exploration of the uncanny valley from the viewpoint of the robot's nonverbal

behaviour. Int J Soc Robot Adv Online Publ 13:1443–1455. https://doi.org/10.1007/s12369-020-00726-w

27. Sato W, Namba S, Yang D, Nishida S, Ishi C, Minato T (2021) An android for emotional interaction: spatiotemporal validation of its facial expressions. Front Psychol 12:800657. https://doi.org/10.3389/fpsyg.2021.800657

28. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791. https://doi.org/10.1038/44565

29. Namba S, Matsui H, Zloteanu M (2021) Distinct temporal features of genuine and deliberate facial expressions of surprise. Sci Rep 11:3362. https://doi.org/10.1038/s41598-021-83077-4

30. Namba S, Nakamura K, Watanabe K (2022) The spatio-temporal features of perceived-as-genuine and deliberate expressions. PLOS ONE 17:e0271047. https://doi.org/10.1371/journal.pone.0271047

31. Perusquía-Hernández M, Dollack F, Tan CK, Namba S, Ayabe-Kanamura S, Suzuki K (2021) Smile action unit detection from distal wearable electromyography and computer vision. In: Vol. 2021 16th IEEE international conference on automatic face and gesture recognition (FG 2021). IEEE Publications, 1–8. https://doi.org/10.1109/FG52635.2021.9667047

32. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191. https://doi.org/10.3758/bf03193146

33. Cohen J (1988) Statistical power analysis for the behavioral science, 2nd edn. Lawrence Erlbaum, Hillsdale, NJ

34. Baltrušaitis T, Mahmoud M, Robinson P (2015) Cross-dataset learning and person-specific normalisation for automatic action unit detection. In: 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), Vol. 6. IEEE Publications, 1–6. https://doi.org/10.1109/FG.2015.7284869

35. Baltrušaitis T, Zadeh A, Lim YC, Morency LP (2018) Openface 2.0: facial behavior analysis toolkit. In: 3th IEEE international conference on automatic face & gesture recognition (FG 2018). IEEE Publications, 59–66. https://doi.org/10.1109/FG.2018.00019

36. Ekman P, Friesen WV, Hager JC (2002) Facial action coding system, 2nd edn. Research nexus e-book, Salt Lake City

37. Ekman P, Rosenberg E (2005) What the face reveals, 2nd edn. Oxford University Press, New York

38. Wood E, Baltruaitis T, Zhang X, Sugano Y, Robinson P, Bulling A (2015) Rendering of eyes for eye-shape registration and gaze estimation. In: proceedings of the IEEE international conference on computer vision, 3756–3764. https://doi.org/10.1109/ICCV.2015.428

39. Brunet JP, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 101:4164–4169. https://doi.org/10.1073/pnas.0308531101

40. Kim H, Park H (2007) Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics 23:1495–1502. https://doi.org/10.1093/bioinformatics/btm134

41. Hutchins LN, Murphy SM, Singh P, Graber JH (2008) Position-dependent motif characterization using non-negative matrix factorization. Bioinformatics 24:2684–2690. https://doi.org/10.1093/bioinformatics/btn526

42. Del Re AC (2013). Compute. es: compute effect sizes. R package version, 0–2

43. Gaujoux R, Seoighe C (2010) A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11:367. https://doi.org/10.1186/1471-2105-11-367

44. Iseki R (2023) Retrieved [Jul 10, 2023] from @@@@

45. Wickham H, Averick M, Bryan J et al (2019) Welcome to the tidyverse. J Open Source Softw 4:1686. https://doi.org/10.21105/joss.01686

46. Namba S, Sato W, Osumi M, Shimokawa K (2021) Assessing automated facial action unit detection systems for analyzing cross-domain facial expression databases. Sensors (Basel) 21:4222. https://doi.org/10.3390/s21124222

47. Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD (2019) Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. Psychol Sci Public Interest 20:1–68. https://doi.org/10.1177/1529100619832930

48. Barrett LF, Mesquita B, Gendron M (2011) Context in emotion perception. Curr Dir Psychol Sci 20:286–290. https://doi.org/10.1177/0963721411422522

49. Chen Z, Whitney D (2019) Tracking the affective state of unseen persons. Proc Natl Acad Sci U S A 116:7559–7564. https://doi.org/10.1073/pnas.1812250116

50. Majima Y, Nishiyama K, Nishihara A, Hata R (2017) Conducting online behavioral research using crowdsourcing services in Japan. Front Psychol 8:378. https://doi.org/10.3389/fpsyg.2017.00378

51. Dawel A, Wright L, Irons J, Dumbleton R, Palermo R, O'Kearney R, McKone E (2017) Perceived emotion genuineness: normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. Behav Res Methods 49:1539–1562. https://doi.org/10.3758/s13428-016-0813-2

52. Broadbent E, Kumar V, Li X, Sollers J, Stafford RQ, MacDonald BA, Wegner DM (2013) Robots with display screens: a robot with a humanlike face display is perceived to have more mind and a better personality. PLOS ONE. https://doi.org/10.1371/journal.pone.0072589

53. Mende M, Scott ML, van Doorn J, Grewal D, Shanks I (2019) Service robots rising: how humanoid robots influence service experiences and elicit compensatory consumer responses. J Mark Res 56:535–556. https://doi.org/10.1177/0022243718822827

54. Cheshin A, Amit A, Van Kleef GA (2018) The interpersonal effects of emotion intensity in customer service: perceived appropriateness and authenticity of attendants' emotional displays shape customer trust and satisfaction. Organ Behav Hum Decis Processes 144:97–111. https://doi.org/10.1016/j.obhdp.2017.10.002

55. Churchill GA, Peter JP (1984) Research design effects on the reliability of rating scales: a meta-analysis. J Mark Res 21:360–375. https://doi.org/10.1177/002224378402100402

56. Preston CC, Colman AM (2000) Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol 104:1–15. https://doi.org/10.1016/s0001-6918(99)00050-5

57. Krosnick JA, Presser S (2010) Question and questionnaire design. In: Marsden PV, Wright JD (eds) Handbook of survey research 2nd. Emerald Publishing Group, Bingley, UK, pp 263–313

58. Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed effects models. J Stat Soft 82:1–26

59. Revelle W, Revelle MW (2015) Package "psych". The Compr R Arch Netw 337

60. Ambadar Z, Cohn JF, Reed LI (2009) All smiles are not created equal: morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. J Nonverbal Behav 33:17–34. https://doi.org/10.1007/s10919-008-0059-5

61. Krumhuber EG, Kappas A (2022) More what Duchenne smiles do, less what they express. Perspect Psychol Sci 17:1566–1575. https://doi.org/10.1177/17456916211071083

62. Scherer KR, Grandjean D (2008) Facial expressions allow inference of both emotions and their components. Cogn Emot 22:789–801. https://doi.org/10.1080/02699930701516791

63. Horstmann G (2003) What do facial expressions convey: feeling states, behavioral intentions, or action requests? Emotion 3:150–166. https://doi.org/10.1037/1528-3542.3.2.150

64. Hui MK, Zhou L (1996) How does waiting duration information influence customers' reactions to waiting for services? J Appl Soc Psychol 26(19):1702–1717

65. Myers BA (1985) The importance of percent-done progress indicators for computer-human interfaces. ACM SIGCHI Bulletin 16(4):11–17

66. Branaghan RJ, Sanchez CA (2009) Feedback preferences and impressions of waiting. Hum factors 51(4):528–538

67. Wintersberger P, Klotz T, Riener A (2020, October) Tell me more: transparency and time-fillers to optimize chatbots' waiting time experience. In: proceedings of the 11th Nordic conference on human-computer interaction: shaping experiences, Shaping society. pp. 1–6

68. de Leeuw JR, Gilbert RA, Luchterhandt B (2023) jsPsych: enabling an open-source collaborative ecosystem of behavioral experiments. J Open Source Softw 8(85):5351. https://doi.org/10.21105/joss.05351

69. Ten Bosch L, Oostdijk N, De Ruiter JP (2004) Turn-taking in social talk dialogues: temporal, formal and functional aspects. In: 9th international conference speech and computer (SPECOM'2004),pp 454–461

70. Sosnowski S, Bittermann A, Kuhnlenz K, Buss M (2006) (2006) Design and evaluation of emotion-display eddie. Intelligent robots and systems. IEEE/RSJ international conference on, IEEE, pp 3113–3118

71. Fukuda T, Taguri J, Arai F, Nakashima M, Tachibana D, Hasegawa Y (2002). Facial expression of robot face for human-robot mutual communication. In: proceedings 2002 ieee international conference on robotics and automation, pp 46–51

72. Glas DF, Minato T, Ishi CT, Kawahara T, Ishiguro H (2016) Erica: the erato intelligent conversational android. In: 2016 25th IEEE International symposium on robot and human interactive communication (RO-MAN), pp 22–29

73. Hegel F, Eyssel F, Wrede B (2010) The social robot 'flobi': keyconcepts of industrial design. In: 19th international symposium inrobot and human interactive communication. IEEE, pp 107–112

74. Mirnig N, Strasser E, Weiss A, Kühnlenz B, Wollherr D, Tscheligi M (2015) Can you read my face? a methodological variation for assessing facial expressions of robotic heads. Int J Soc Robot 7:63–76

75. Hömke P, Holler J, Levinson SC (2017) Eye blinking as addressee feedback in face-to-face conversation. Res Lang Soc Interact 50:54–70. https://doi.org/10.1080/08351813.2017.1262143

76. Hömke P, Holler J, Levinson SC (2018) Eye blinks are perceived as communicative signals in human face-to-face interaction. PLOS ONE 13:e0208030. https://doi.org/10.1371/journal.pone.0208030

77. Elfenbein HA (2013) Nonverbal dialects and accents in facial expressions of emotion. Emot Rev 5:90–96. https://doi.org/10.1177/1754073912451332

78. Cordaro DT, Sun R, Keltner D, Kamble S, Huddar N, McNeil G (2018) Universals and cultural variations in 22 emotional expressions across five cultures. Emotion 18:75–93. https://doi.org/10.1037/emo0000302

79. Fang X, Sauter DA, Heerdink MW, van Kleef GA (2022) Culture shapes the distinctiveness of posed and spontaneous facial expressions of anger and disgust. J Cross Cult Psychol 53:471–487. https://doi.org/10.1177/00220221221095208

80. Jack RE, Sun W, Delis I, Garrod OG, Schyns PG (2016) Four not six: revealing culturally common facial expressions of emotion. J Exp Psychol Gen 145:708–730. https://doi.org/10.1037/xge0000162

81. Masuda T, Ellsworth PC, Mesquita B, Leu J, Tanida S, Van de Veerdonk E (2008) Placing the face in context: cultural differences in the perception of facial emotion. J Pers Soc Psychol 94:365–381. https://doi.org/10.1037/0022-3514.94.3.365

82. Namba S, Rychlowska M, Orlowska A, Aviezer H, Krumhuber EG (2020) Social context and culture influence judgments of non-Duchenne smiles. J Cult Cogn Sci 4:309–321. https://doi.org/10.1007/s41809-020-00066-1

83. Parkinson B (2017) Interpersonal effects and functions of facial activity. In: Fernández-Dols JM, Russell JA (eds) The science of facial expression. Oxford University Press, New York, pp 435–456

84. Kiilavuori H, Sariola V, Peltola MJ, Hietanen JK (2021) Making eye contact with a robot: psychophysiological responses to eye contact with a human and with a humanoid robot. Biol Psychol 158:107989

85. Krumhuber EG, Kappas A, Manstead ASR (2013) Effects of dynamic aspects of facial expressions: a review. Emot Rev 5:41–46. https://doi.org/10.1177/1754073912451349

86. Krumhuber EG, Skora LI, Hill HCH, Lander K (2023) The role of facial movements in emotion recognition. Nat Rev Psychol 2:283–296. https://doi.org/10.1038/s44159-023-00172-1

87. Sato W, Krumhuber EG, Jellema T, Williams JHG (2019) Editorial: dynamic emotional communication. Front Psychol 10:2836. https://doi.org/10.3389/fpsyg.2019.02836

88. Ertugrul IO, Cohn JF, Jeni LA, Zhang Z, Yin L, Ji Q (2020) Crossing domains for au coding: perspectives, approaches, and measures. IEEE Trans Biom Behav Identity Sci 2:158–171. https://doi.org/10.1109/tbiom.2020.2977225

89. Cohn JF, Ertugrul IO, Chu WS, Girard JM, Jeni LA, Hammal Z (2019) Affective facial computing: generalizability across domains, Multimodal behavior analysis in the wild. Academic Press, USA, pp 407–441

Shushi Namba received a PhD degree from Hiroshima University, Japan. In October 2020, he was a researcher at the Guardian Robot Project, RIKEN. Since April 2023, he has been working as an associate professor at Hiroshima University. He is also a visiting researcher of the Guardian Robot Project, RIKEN. His research interests include facial expressions and emotional communication.

Wataru Sato is the Team Leader of the Psychological Process Research Team at the Guardian Robot Project, RIKEN. His research interests include the psychology and neuroscience of social interactions. He received his B.S., M.S., and Ph.D. degrees from Kyoto University, Japan.

Saori Namba is the Research assistant at Hiroshima University. Her research interests include clinical psychology. She received M.S. degrees from Hiroshima University, Japan.

Alexander Diel completed his B.Sc and M.Sc in Psychology at the Ruhr-University Bochum in Germany. He did his Ph.D on the uncanny valley in Cardiff University and was a visiting researcher at the RIKEN Institute in Kyoto, Japan. He is currently a postdoctoral researcher on digital mental health at the LVR University Essen in Germany where he leads the work unit PsychosomaTech.

Carlos Toshinori Ishi received the PhD degree in Engineering from The University of Tokyo, Japan, in 2001. He worked at the JST/CREST Expressive Speech Processing Project from 2002 to 2004 at ATR. He joined ATR Intelligent Robotics and Communication Labs, since 2005, and is the group leader of the Department of Sound Environment Intelligence at ATR Hiroshi Ishiguro Labs, since 2013. He joined RIKEN Guardian Robot Project from 2021. His research interests include speech processing, paralinguistic analysis, and social robots.

Takashi Minato obtained a Ph.D. degree in Engineering from Osaka University in 2004. In December 2001, he was a researcher at CREST, JST and worked as an assistant professor of the Department of Adaptive Machine Systems, Osaka University from September 2002 until 2006. In June 2006, he began to work as a researcher for ERATO, JST. In January 2011, he was a researcher in the ATR Hiroshi Ishiguro Laboratories. Since May 2020, he has been working as a team leader in Guardian Robot Project, RIKEN. He is also a visiting researcher of the ATR. His research interests include human-robot interaction and android science.