In the Shades of the Uncanny Valley: An Experimental Study of Human–Chatbot Interaction

Article *in* Future Generation Computer Systems · February 2018

DOI: 10.1016/j.future.2018.01.055

CITATIONS
231

READS
6,591

4 authors:

Leon Ciechanowski
SWPS University of Social Sciences and Humanities
16 PUBLICATIONS **413** CITATIONS
SEE PROFILE

Aleksandra Przegalinska
Harvard University
42 PUBLICATIONS **517** CITATIONS
SEE PROFILE

Mikołaj Magnuski
SWPS University of Social Sciences and Humanities
16 PUBLICATIONS **539** CITATIONS
SEE PROFILE

Peter Gloor
Massachusetts Institute of Technology
319 PUBLICATIONS **5,228** CITATIONS
SEE PROFILE

In the Shades of the Uncanny Valley: An Experimental Study of Human–Chatbot Interaction

Leon Ciechanowski[a1], Aleksandra Przegalinska[b], Mikolaj Magnuski[a], Peter Gloor[c]

[a] Department of Psychology, University of Social Sciences and Humanities, Chodakowska 19/31, Warsaw, Poland.

[b] Department of Management, Kozminski University, Jagiellonska 57/59 Warsaw, Poland.

[c] MIT Center for Collective Intelligence, 245 First Street, E94-1509 Cambridge, MA, USA.

Abstract: This project has been carried out in the context of recent major developments in botics and more widespread usage of virtual agents in personal and professional sphere. The general purpose of the experiment was to thoroughly examine the character of the human–non-human interaction process. Thus, in the paper, we present a study of human–chatbot interaction, focusing on the affective responses of users to different types of interfaces with which they interact. The experiment consisted of two parts: measurement of psychophysiological reactions of chatbot users and a detailed questionnaire that focused on assessing interactions and willingness to collaborate with a bot. In the first quantitative stage, participants interacted with a chatbot, either with a simple text chatbot (control group) or an avatar reading its responses in addition to only presenting them on the screen (experimental group. We gathered the following psychophysiological data from participants: electromyography (EMG), respirometer (RSP), electrocardiography (ECG), and electrodermal activity (EDA). In the last, declarative stage, participants filled out a series of questionnaires related to the experience of interacting with (chat)bots and to the overall human–(chat)bot collaboration assessment. The theory of planned behavior survey investigated attitude towards cooperation with chatbots in the future. The social presence survey checked how much the chatbot was considered to be a "real" person. The anthropomorphism scale measured the extent to which the chatbot seems humanlike. Our particular focus was on the so-called uncanny valley effect, consisting of the feeling of eeriness and discomfort towards a given medium or technology that frequently appears in various kinds of human–machine interactions. Our results show that participants were experiencing lesser

[1] Corresponding author

E-mail addresses: lciechanowski@swps.edu.pl (L. Ciechanowski), aprzegalinska@kozminski.edu.pl (A. Przegalinska), mmagnuski@swps.edu.pl (M. Magnuski), pgloor@mit.edu (P. Gloor).

L. Ciechanowski, A. Przegalinska, and M. Magnuski equally contributed to the article.

uncanny effects and less negative affect in cooperation with a simpler text chatbot than with the more complex, animated avatar chatbot. The simple chatbot have also induced less intense psychophysiological reactions. Despite major developments in botics, the user's affective responses towards bots have frequently been neglected. In our view, understanding the user's side may be crucial for designing better chatbots in the future and, thus, can contribute to advancing the field of human–computer interaction.

Highlights

- A two-stage experiment focusing on human–chatbot interaction was conducted.
- Methodology consisted of processing psychophysiological data and collecting declarative responses through questionnaires.
- Major findings confirm stronger negative affect, emotional arousal, and increased uncanny valley effect ("weirdness" or discomfort) in the chatbot enriched with animated avatar and sound.
- Participants declared interest in cooperation with both types of chatbots in the future.

Keywords: Human–Computer Interaction, Chatbots, Affective Computing, Psychophysiology, Uncanny Valley

1. Introduction

Human–computer interaction (HCI) is currently an area of singular importance, and a key to understanding it is an appreciation of the fact that interactive interfaces mediate the redistribution of cognitive tasks between humans and machines. Chatbots are an interesting case in HCI, as they are designed to interact with users through natural languages. This technology, started in the 1960s, initially aimed to determine whether chatbot systems could fool users into believing that they were real humans. Quite early, it was confirmed that cheating was possible; nonetheless, the widespread usage of chatbots remained obscure for several decades. Currently, however, chatbot systems are built not only to mimic human conversation and entertain users but are also for use in applications in education, information retrieval, business, and e-commerce. In fact, chatbots are a perfect example of the implementation of state-of-the-art consumer-oriented artificial intelligence that simulates human behaviour based on formal models, and they have been an interesting subject for the research of patterns of human and non-human interaction as well as issues related to assigning social roles to others, finding patterns of successful and unsuccessful interactions, and establishing social relationships and bonds.

Being designed to aid cognition and simplify tasks, interfaces function as specific cognitive artefacts that provide the most flexible representational medium we have ever known and enable novel forms of communication. Several spheres of research interest in HCI exist. The first of these explores interfaces that expand representational possibilities beyond metaphors to which

people have already become accustomed, such as icons on a desktop [1]. A second sphere of interest is related to technology-supported cooperative work [2]. A third draws on what we know about human perception and cognition, coupling it with the task analysis method. The latter research field is crucial for our experiment.

To date, only a few controlled experiments have been conducted to directly examine the interaction between humans and chatbots [3–5]. At the same time, almost no study has implemented psychophysiological instruments for the research of the human–chatbot interaction, although in the paradigm of social interaction—or social cognition—a large number of studies have come up with various psychophysiological and neuroimaging methods (see [6]), especially the meta-analysis of 70 social cognition studies using functional neuroimaging techniques [7]. Similar relationships—albeit in virtual reality [8,9]—have tracked social interactions in business and everyday situations in virtual reality as well as their psychophysiological or neural correlates. In addition, Hofree et al. [10] examined human-android interaction using EMG but focused on simple emotional responses, not full verbal interaction.

We place our research in the context of the ongoing process of introducing artificial intelligence in the area of social interaction with people, with particular emphasis on interactions in the professional sphere (e.g., business and strategic management). In doing so, we argue for paying greater attention to the factors that cause either trust or resistance towards such technological innovations in professional and social life. Our work, which focuses on the user and his/her declarative and psychophysiological responses to a bot, will fill a gap in the HCI research, where little attention has thus far been paid to the socio-cognitive nature of the interaction between man and technology in general and chatbots in particular.

In order to understand and better explain people's complex relationships with information technology in the professional environment, our study emphasises the user. Instead of focusing on building and perfecting yet another technology, we concentrate on the other side of the "attention economy"—namely, technology's interlocutors (i.e., humans). Essentially, we seek to compare the way people interact with different chatbots that imitate both human ways of interacting and human expertise.

1.1. Chatbots and their application and role in society

Just as people use language for human communication, people want to use their language to communicate with computers. Kacprzyk and Zadrozny [11] argued that the best way to facilitate HCI is by allowing users "to express their interest, wishes, or queries directly and naturally, by speaking, typing, and pointing". Morrissey and Kirakowski [12] made a similar point in their criteria for development of a more human-like chatbot.

This was the motivation behind the development of chatbots. A chatbot system is a software program that interacts with users using natural language. Different terms have been used for a chatbot, such as machine conversation system, virtual agent, dialogue system, and chatterbot. Initially, developers built and used chatbots for fun and used simple keyword matching techniques to find a match for a user input, such as ELIZA [13,14]. Before the arrival of graphical user interfaces, the seventies and eighties saw rapid growth in text and natural language interface research [15]. Since that time, a range of new chatbot architectures have been

developed, such as MegaHAL [16], CONVERSE [17], and A.L.I.C.E. [18], which was first implemented by Wallace in 1995. A.L.I.C.E.'s knowledge about English conversation patterns is stored in artificial intelligence markup language (AIML) files. AIML is a derivative of extensible markup language (XML) that was developed by Wallace. The Alicebot free software community from 1995 onwards enabled people to input dialogue pattern knowledge into chatbots based on the A.L.I.C.E. open-source software technology.

Currently, the choice of chatbots is much broader and includes a great deal of machine learning-supported consumer technologies, such as Siri and Cortana [19,20]. They may take the shape of virtual agents or physical objects, such as Alexa [21] or Jibo [22,23], which can also be researched from the perspective of the proxemic relations they maintain with the users, as well as gestural communication. In the study we present here, we have used a simple state-of-the-art virtual assistant integrated with a help desk and supported by supervised learning and intelligent dialog management software. Our bot, "Ola", was based on a stochastic data model with synonyms and meaning detection, using a customized engine written in PHP and Node.js (Javascript). It was designed to provide detailed information about a Central European university (Kozminski University) as a dedicated service for current and future students.

1.2. Previous research on the uncanny valley

Today, we know that many kinds of media are capable of eliciting, to varying degrees, different kinds of social responses, including verbal [5], gestural, and visual responses [24]. However, to date, only a few experiments have directly examined communication between humans and chatbots. Wide problematics of human attitudes towards humanoid technologies were first discussed in Mori's [25] uncanny valley hypothesis, which predicted that perceptual difficulty in distinguishing between a human-like object and its human counterpart will evoke negative affects. The hypothesis suggests that humanoid objects which appear almost, but not exactly, like real human beings elicit uncanny, or strangely familiar, feelings of revulsion or eeriness in observers. In fact, Mori observed a heightened sensitivity to defects in near-humanlike forms—an "uncanny valley" in what is otherwise a positive relationship between human likeness and familiarity. Since then, research has focused on affect, but still fairly little is known about how humanoid objects are actually perceived.

Reasons for the appearance of the uncanny valley effect have not been fully explained [26,27]. One explanation could be mate selection, where uncanny stimuli can elicit aversion by activating an evolved cognitive mechanism for the avoidance of selecting mates with low fertility and poor hormonal health—which can possibly be detected subconsciously by observing face features [28]. Another reason is pathogen avoidance, where uncanny stimuli may activate a cognitive mechanism that originally evolved to motivate the avoidance of potential sources of pathogens by eliciting a disgust response. The negative effect associated with uncanny stimuli is produced by the activation of conflicting cognitive representations and may seem to pose a threat to humans' distinctiveness and identity. Negative reactions towards very humanlike robots can be related to the challenge that these kinds of robots lead to the categorical human—namely, non-human distinction.

Also, entities with human and nonhuman traits undermine our sense of human identity by linking qualitatively different categories, human and nonhuman, by a quantitative metric: degree of human likeness. Aside from more biological and evolutionary psychology responses, explanations have been related to cultural and religious norms. The uncanny valley may be symptomatic of entities that "elicit a model of a human other but do not measure up to it" [29] as well as a religious definition of human identity. Some view the existence of artificial but humanlike entities as a threat to the concept of human identity.

One recent studies [30] hypothesized that a human-looking android may be perceived as uncanny because it elicits the fear of death. The author attempted to verify this through questions designed to measure distal terror management defences, such as worldview protection. The results were rather favourable, as the group exposed to an image of an uncanny robot consistently preferred information sources that supported their worldview relative to the control group. This, however, has only applied to one particular stimulus, so it was quite difficult to generalize across stimuli. In the experiment, where participants were presented with short video clips of a wide range of mainly android and humanoid robots engaged in various activities in different settings, the results did not indicate a single uncanny valley effect for a particular range of human likeness. Yet another experiment [29] in which the android's responses were identical showed that human responses to different androids varied according to their beliefs.

Hanson [31] noticed that the notion that the uncanny valley can be escaped through varying factors unrelated to human likeness. Although Hanson found that morphing from a mechanical looking robot to an android produced an uncanny valley on both a familiarity scale and an appealing scale, as well as a peak in an eeriness scale, these effects were greatly reduced by tuning the morphs.

A recent work on artificial gaze [32] aimed at determining how human and robot gaze can influence the speed and accuracy of human action. Another more recent experiment focusing on speech shadowing [33] created situations in which people conversed in person with a human whose words were determined by a conversational agent computer program. The method involved a person repeating vocal stimuli originating from a separate communication source in real time. Humans shadowing conversational agent sources (bots) thus became hybrid agents ("echo-borgs") capable of face-to-face interlocution. In a series of three studies, the authors noticed that, unlike those who engaged with a text interface, the vast majority of participants who engaged with an echo-borg did not sense any robotic interaction. These findings have implications for HCI as they show that the human body, as a channel of communication, fundamentally alters the social and psychological dynamics of interactions with machine intelligence.

Finally, our main reference study [34] investigated the effect of self-involvement during social interaction on attention, arousal, and facial expression. Specifically, the study sought to disentangle the effect of being personally addressed from the effect of decoding the meaning of another person's facial expression. To this end, eye movements, pupil size, and facial electromyographic (EMG) activity were recorded while participants observed virtual characters looking at either them or someone else. In dynamic animations, the virtual characters then displayed either socially relevant facial expressions (similar to those used in everyday life situations) to establish interpersonal contact or arbitrary facial movements. The results show that attention allocation, as assessed by eye-tracking measurements, was specifically related to self-

involvement, regardless of the social meaning being conveyed. Arousal, as measured by pupil size, was primarily related to the perception of the virtual character's gender. In contrast, facial EMG activity was determined by the perception of socially relevant facial expressions, regardless of whether the virtual characters were looking at the participants or at someone else.

2. Hypotheses

The main hypothesis of the study was that significant differences in both psychophysiological and declarative responses of study participants would be revealed in relation to the type of bot with which the participants interacted. These differences were expected to depend on the type of the entity with which the tested groups were interacting (i.e., avatar versus text chatbot). We assumed that emotional and physiological responses would be most intense in the interaction with the more human-like avatar chatbot but reduced in the interaction with the simpler text chatbot. We also hypothesised that the avatar chatbot would induce a stronger uncanny valley effect than the text chatbot.

3. Methodology and experiment design

In order to investigate potential differences among various types of chatbots as well as the influence on their users, we carried out a two-stage experiment. In the first quantitative stage, participants interacted with a chatbot, either with a simple text chatbot (control group) referred to here as TEXT (Fig. 1) or an avatar reading its responses in addition to only presenting them on the screen (experimental group), which we call AVATAR (Fig. 2). Both chatbots were designed to be Kozminski Academy website assistants; therefore, they mainly provided information about the enrolment process for new students. Nevertheless, they were capable of casual conversation.

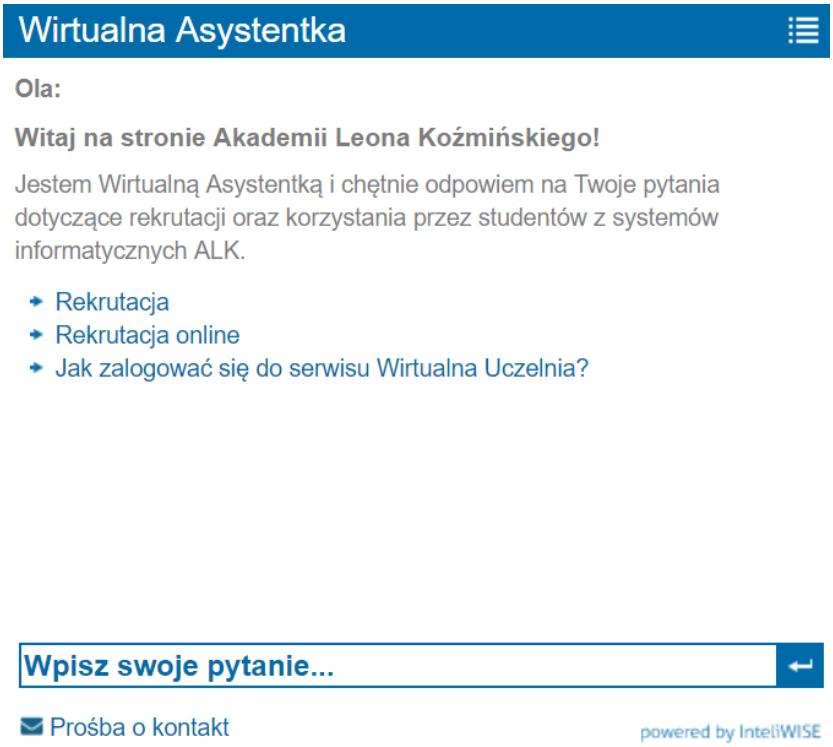

Figure 1. Screenshot of the TEXT chatbot that participants interacted with. It is a "virtual assistant" ["Wirtualna Asystentka"] named "Ola". Additional text on the screen are as follows: "Witam na stronie Akademii Leona Koźmińskiego!" ["Welcome to the Leon Koźmiński's Academy"]; "Jestem Wirtualną Asystentką i chętnie odpowiem na Twoje pytania dotyczące rekrutacji oraz korzystania przez studentów z systemów informatycznych ALK." ["I am a Virtual Assistant and I am happy to answer your questions about the enrollment process and the use of ALK IT systems."]; "Rekrutacja; Rekrutacja online; Jak zalogować się do serwisu Wirtualna Uczelnia?" ["Enrollment; Enrollment online; How to log in to the Virtual University website?"]; "Wpisz swoje pytanie…" ["Write your question here…"]; "Prośba o kontakt" ["I wish to contact you"] - this hyperlink was inactive.

The chatbot's and participants' responses were presented one after another (in a chat form) on the screen. Participants added their input in the field "Write your question here…" ["Wpisz swoje pytanie…"].

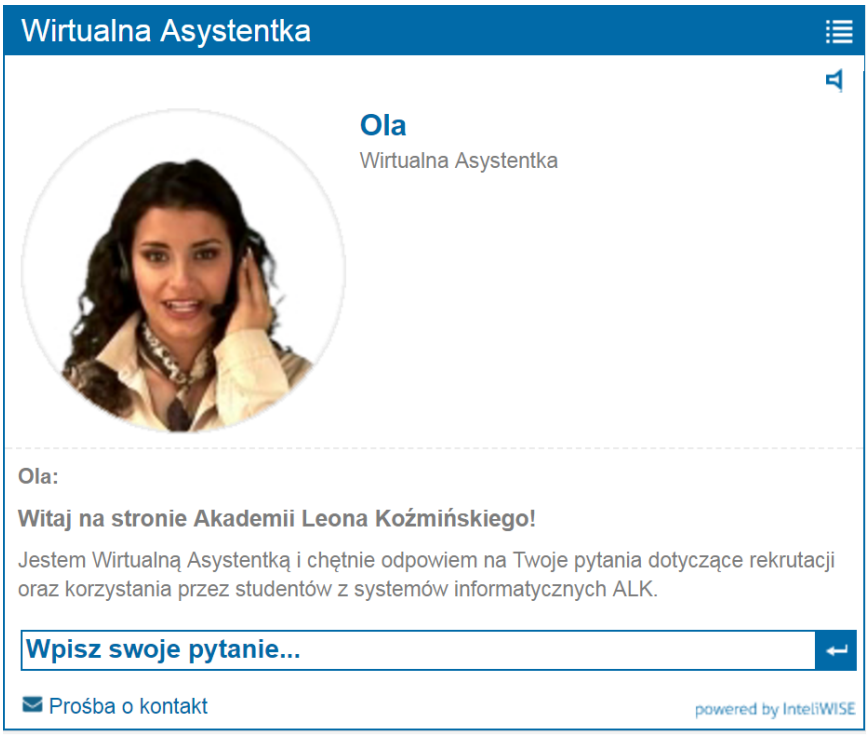

Figure 2. Screenshot of the AVATAR chatbot participants used. The visible avatar was animated (moved according to speech produced on the basis of its text responses presented on the screen).

Figure 3 presents the methodological scheme. After participants were randomly assigned to one of the groups, we placed electrodes used in psychophysiological analyses (more information is provided in Section 3.2. Psychophysiological measures) and the interaction with the chatbot began. Each participant was left alone in the lab room and was instructed to discuss the academy enrolment process and ask the chatbot about six specific aspects regarding this topic. Such formulation of universal tasks across both groups allowed us to achieve at least some level of experimental control over the process involving participants. After completing the task, participants were supposed to engage in casual conversation with the chatbot. Participants talked with the chatbot by writing messages/questions on a computer keyboard. Upon pressing "ENTER", the chatbot's response appeared on the screen with a maximum one-second delay. Afterwards, participants took a test of knowledge based on what they learned from the chatbot (this step played the role of a filler task). In the end, participants filled out a few questionnaires (for more details, see this Section below).

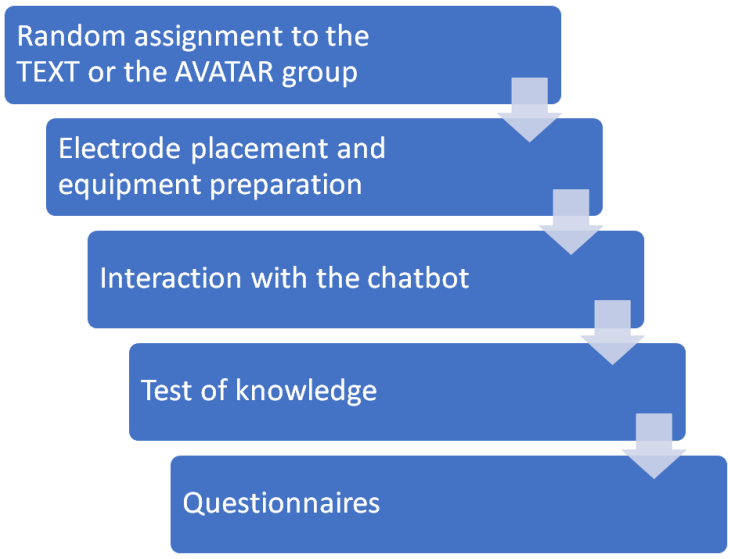

Figure 3. Scheme of the research protocol.

We gathered the following psychophysiological data from participants on an ongoing basis throughout the entire interaction with the chatbot:

- EMG data were collected from eyebrow-wrinkling muscles (*musculus corrugator supercilii*) and smiling muscles (*musculus zygomaticus*), in accordance with standard guidelines [35]. This measure was mainly used to measure emotional arousal—in particular, anger, happiness, fear, and sadness.
- –Respirometer (RSP) is an auxiliary measure used in ECG analysis.
- –Electrocardiogram (ECG) measures the heart rate, which is correlated with the activity of a sympathetic nervous system, which may trigger the "fight-or-flight" response [36], also called the acute stress response. This physiological reaction occurs in response to a perceived harmful event, attack, or other stimulus causing distress or anxiety [36].
- Electrodermal activity (EDA) is a measure of sweat gland permeability, observed as changes in the electric resistance of the skin. It may be triggered by general emotional arousal, fear, or surprise.

In the last, qualitative stage, participants filled out a series of questionnaires related to the experience of interacting with (chat)bots and to the overall human–(chat)bot collaboration assessment [37–39]:

- The theory of planned behaviour survey investigated attitude towards cooperation with chatbots in the future.
- The social presence survey checked how much the chatbot was considered to be a "real" person.
- The anthropomorphism scale measured the extent to which the chatbot seems humanlike.

Each questionnaire consisted of questions that later became items in more general factors, such as uncanny valley and competence.

In the analysis stage, the questionnaire data were included in a multivariate analysis of variance in the mixed scheme, where the assessment of characteristic parameters of chatbots was a within-

subject variable (measured with questionnaires and factors created on the basis of these) whereas the between-subject variable consisted of the type of entity present in interactions.

Psychophysiological measurements were conducted with the BIOPAC MP150 system using modules for EMG, ECG, RSP, and EDA. The EMG was recorded from the *zygomaticus major* ("smile muscle") and *corrugator supercilii* ("frown muscle"), and EDA was recorded with electrodes placed on the small and ring finger of the nondominant hand. The data were recorded using AcqKnowledge software with a sampling rate of 2000 Hz. The chatbot's responses were synchronised with the BIOPAC device using custom Python code that captured the screen every 10 ms and detected the moment when the chatbot's textbox was refreshed (this refresh happened only at the chatbot's response onset).

3.1. Participants

Because the study aims to discover the general effects in the domain of human–chatbot interactions that are applicable to all people, we did not foresee any special procedure for the selection of respondents for the groups aside from standard prerequisites like no neurological or psychological disorders and normal or corrected-to-normal vision. We had 31 participants in total (16 in the TEXT group and 15 in the AVATAR group; 18 women). The mean age was 26.9 years (SD = 5.88), 1 person was a primary school graduate, 10 secondary, and 20 had higher education.

3.2. Psychophysiological measures

3.2.1. EDA

According to the psychophysiological literature, the EDA needs time to build up—up to 5 or 6 seconds—after a stimulus onset [40]. Because the stimulus was text/voice and participants needed some time to read it, we decided to use segments extending from 2 seconds before until 7 seconds after each chatbot's response onset. Because these segments are centred with respect to the chatbot's response onset, they are conventionally called "event-related". This event-related EDA (ER-EDA) signal was baseline-corrected with 1 second prior to each chatbot's response. Baseline-correction is commonly used to emphasise changes in the signal evoked by some event of interest (here, the chatbot's response) and is performed by removing the mean of the baseline period from the whole signal. Segments overlapping with the next chatbot's response or when the interval between two consecutive chatbot responses was shorter than 10 seconds were removed, leaving an average of 33.17 segments per participant (SD = 22.00; removed segments: M = 13.33; SD = 20.76). Prior to baseline correction and uniquely for EDA analysis, the ER-EDA signal was *z*-scored (centred and divided by standard deviation). *Z*-scoring the ER-EDA signal was dictated by huge inter-individual differences in the EDA amplitude.

We subsequently averaged all ER-EDA segments for each subject, which resulted in one *z*-scored ER-EDA time series per subject. These time series were compared across experimental conditions with a Welch *t*-test at every time-sample. A Welch *t*-test was chosen over Student *t*-test so that the equal-variance assumption could be dropped (for more arguments for using the

Welch *t*-test by default, see [41]). We also compared ER-EDA variance between the AVATAR and TEXT groups at each time sample using a Levene's test. As the last step of the EDA analysis, we correlated the *z*-scored ER-EDA signal at each time sample with competence and uncanny valley questionnaire factors.

3.2.2. ECG

The ECG signal was analysed using the ECG R-peak detection algorithm implemented in the BioSPPy Python library [42], following a standard approach [43]. After detecting individual R-peaks instantaneous heart rate (IHR) was computed for each pair of R-peaks. This IHR value was assigned to the time occurence of the second of the two R-peaks. This yielded an irregularly sampled signal of IHR which was then linearly interpolated to span all the timesamples of analyzed segments. We used segments ranging from -4 to 7 seconds with respect to each chatbot's response onset and baselined with the average of time range between the -4 to -1 second prior to the chatbot's response; all the following operations (segment selection and averaging) were analogous to what is reported in the EDA analysis section. To obtain one event-related heart rate time series (tachogram) for every participant, all segments were averaged within subjects. This event-related heart rate signal was compared across experimental conditions with a Welch *t*-test.

3.2.3. EMG

We compared differences in facial muscle activity responsible for smiling (*zygomaticus*) and frowning (*corrugator supercilii*) between the AVATAR and TEXT groups. In order to analyse muscle activity evoked by the chatbot's response, the EMG signal was time-locked to each chatbot's response onset (starting from the -2 second before until 7 seconds after) and baselined with the signal preceding the chatbot's response (for a more detailed description, see Section 3.2.1. EDA). EMG muscular activity is manifested in high frequencies ranging from 20 to 200 Hz. We calculated the average frequency content within this range using Morlet wavelets. The 20–200 Hz frequency range was divided into 20 linearly increasing steps, and the power of each frequency was extracted by convolving the signal with 200 ms long Morlet wavelets in the given frequency. Next, all frequency steps within each segment were averaged, giving rise to average high-frequency EMG response timecourse for each segment. Then, to obtain one time-resolved response for each muscle group and participant, all segments were averaged. Finally, these averaged EMG signals were *z*-scored within participants and smoothed using moving average with a 0.5-second window.

4. Results

4.1. Questionnaire results

The questionnaires revealed several interesting facts. First, no differences were observed across gender or education level for each questionnaire factor (all p-values are above .05 for

between groups comparisons). The only significant differences were observed between groups; thus, there was a main effect of the type of chatbot with which the participants interacted. In general, the participants were keen to cooperate with the chatbot, although several of them expressed irritation, or even frustration, during the debriefing session. Nonetheless, both groups had a good opinion about using robots in the workplace, although the opinion was significantly better in the group interacting with the TEXT chatbot, $F(1,29) = 5.452$; $p = .027$ (see Fig. 4). Similarly, the negative affect evaluation factor indicates that participants felt more negative emotions when interacting with the AVATAR chatbot, $F(1,29) = 7.621$; $p = .009$, compared to the TEXT chatbot.

The boxplot in Figure 4 presents the mean results of assessing the nonsupportive anthropomorphic traits factor. Here, we see again that the TEXT chatbot was associated with fewer negative traits, $F(1,29) = 9.224$; $p = .005$. The results also show that the behavioural traits factor was low in both groups, although it was significantly higher for the TEXT chatbot, $F(1,29) = 6.193$; $p = .019$.

Neither of the chatbots was deemed to be particularly competent. The two groups differ significantly, $F(1,29) = 17.023$; $p < .001$.

The uncanny valley factor shows that the TEXT chatbot is considered significantly "less weird" and less inhuman than the AVATAR chatbot, $F(1,29) = 8.309$; $p = .007$.

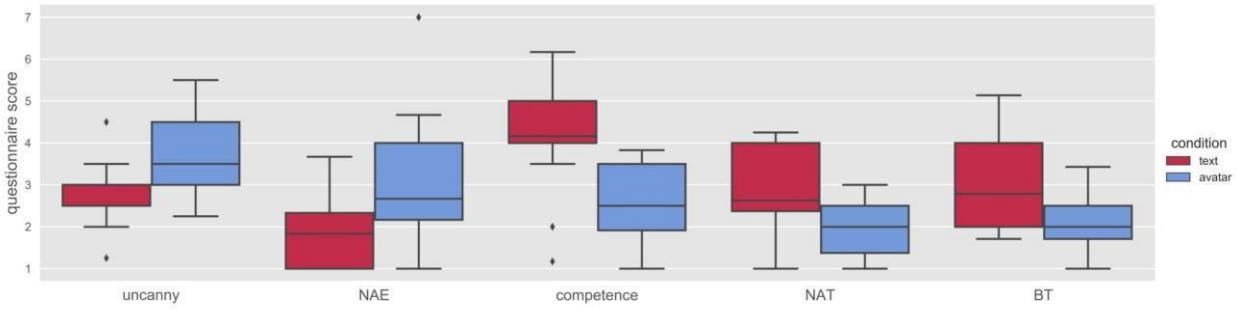

Figure 4. Differences in questionnaire factors' scores between experimental conditions. The factors are: uncanny—uncanny valley measure, NAE—negative affect evaluation, competence—competence dimension, NAT—non-supportive anthropomorphic traits, and BT—behavioural traits. All presented differences are statistically significant (see the main text for details). The higher a factor is on the scale, the greater its intensity.

In order to further investigate the relationships between different dimensions of the chatbots, we have calculated correlations between the uncanny valley factor and the other factors, as this is the aspect of the utmost interest for human–chatbot interaction analyses. The data indicate a strong positive correlation between the uncanny valley factor and negative affect evaluation ($r = .594$, $p < .001$); thus, the more a chatbot was perceived as inhuman or "weird", the more it was disliked. On the contrary, we observed a strong negative correlation between the uncanny valley factor and the dimension of competence ($r = -.654$, $p < .001$); in other words, the more a chatbot was perceived as inhuman, the less competent it seemed to participants (Fig. 5).

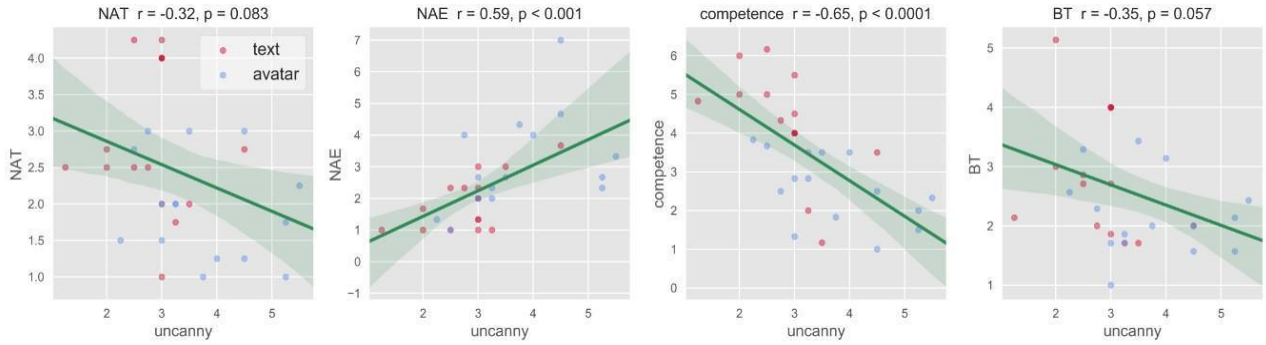

Figure 5. Correlations between the uncanny valley factor and other questionnaire factors.

4.2. Psychophysiological results

4.2.1. EDA results

We observed significant differences in the EDA signal between the TEXT and AVATAR conditions around the time of the chatbot's response onset (before and at time 0) and in the time window ranging from the third to the fourth second after the chatbot's response onset (Fig. 6). The first difference occurs between -0.75 and -0.49 seconds, where the EDA amplitude is higher for the TEXT group (*t* values ranged from $t = 2.07$, $p = 0.048$ to $t = 3.48$, $p = 0.0017$). The second difference (AVATAR > TEXT) ranges from -0.29 to 0.33 seconds (*t* values ranged from $t = -2.05$, $p = .0496$ to $t = -2.94$, $p = .0068$). The third and longest effect occurs in the time window starting at 2.98 and ending at 4.01 seconds after chatbot onset. In this time window, the EDA amplitude is higher for the AVATAR group (*t* values ranged from $t = -2.05$, $p = .0499$ to $t = -2.43$, $p = .0219$).

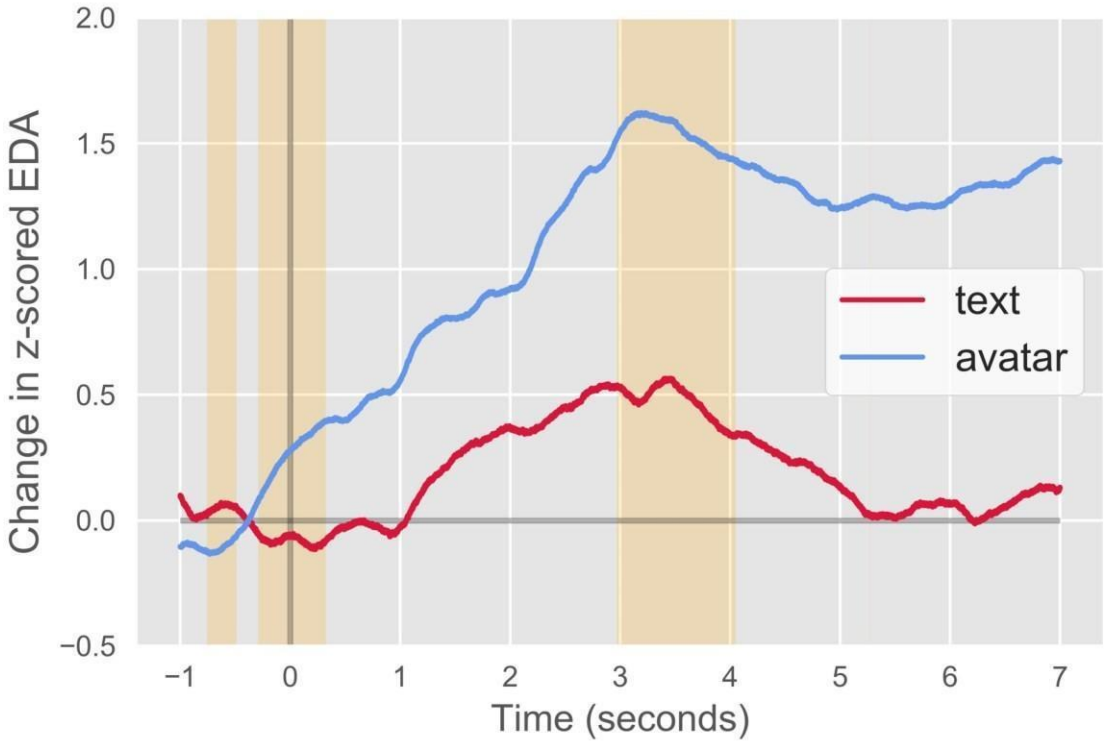

Figure 6. Average standardized baseline-corrected EDA, time-locked to onset of chatbot's response (marker at point 0).

To compare between-group differences in the EDA amplitude variance, we used Levene's test (see Fig. 7). This procedure allowed us to observe significant differences between the two groups in the time window ranging from 3.13 to 4.64 seconds after the chatbot's response onset, where the variance was higher for the TEXT group (Levene W values ranged from $W = 4.21$, $p = .0496$ to $W = 15.89$, $p = .0004$).

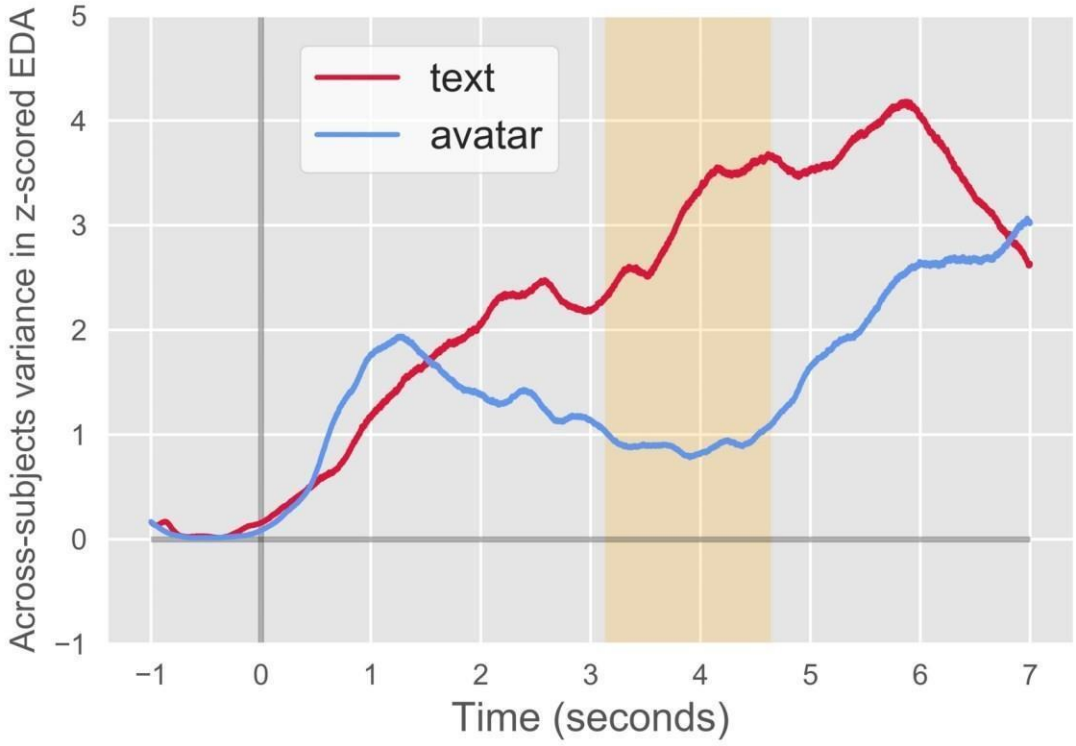

Figure 7. Variance of standardized, baseline-corrected EDA signal with significant differences in variance marked in orange. Time is centred with respect to the moment of the chatbot's response.

4.2.2. ECG results

We first compared average heart rate (BPM) between the TEXT and AVATAR groups using the Welch *t* test. This analysis uncovered a significant between-group difference (t = -2.81; p = .009). Average heart rate was higher in the AVATAR group than in the TEXT group (see Fig. 8).

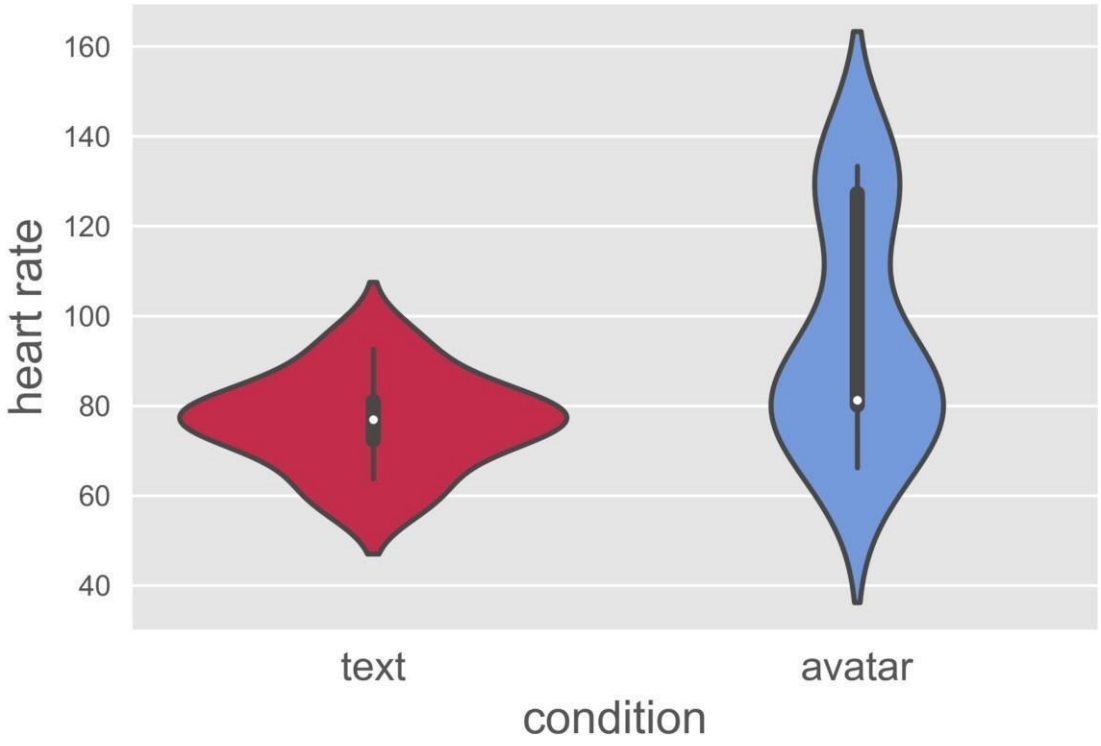

Figure 8. Violin plots of average heart rate (expressed in BPM) in both experimental groups. We observed a significant (p = .002) difference between the groups. Participants interacting with the AVATAR chatbot had a higher average heart rate than participants interacting with the TEXT chatbot.

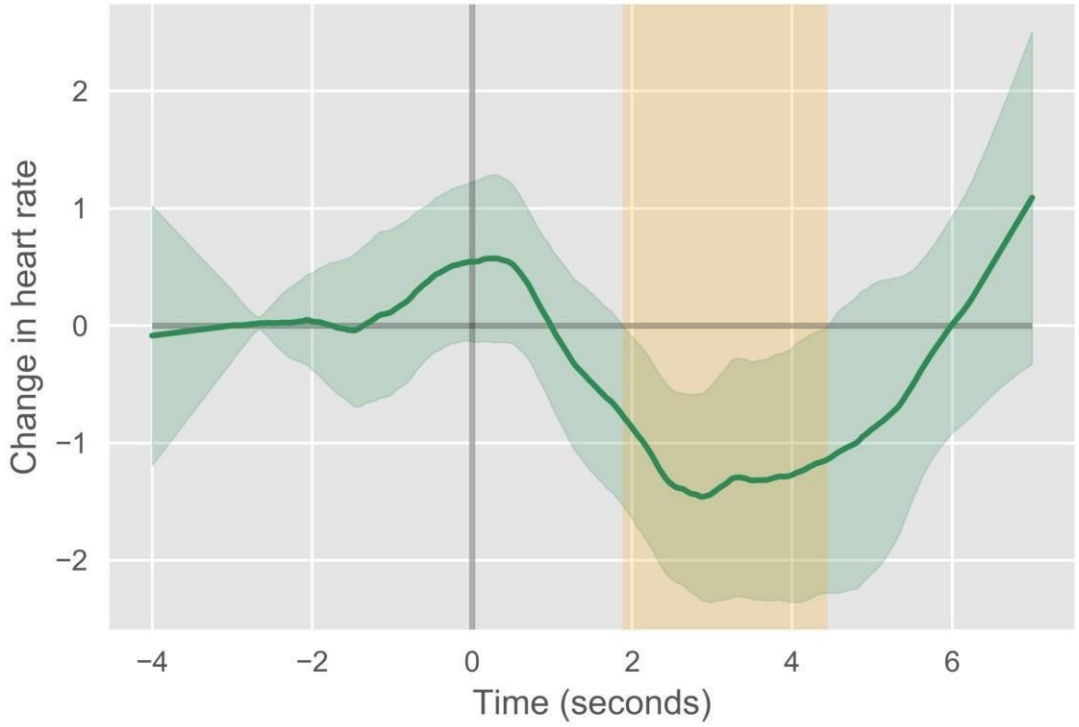

Figure 9. Time-resolved changes in the average heart rate (BPM) relative to baseline (time window from -4 to -1 seconds) in both groups. Time is measured with respect to the onset of the chatbot response. The yellow indicates a significant decrease of heart rate in reference to the baseline.

Next, we conducted a time-resolved analysis by comparing differences in heart rate changes evoked by the chatbot's response. This analysis did not reveal between-group differences ($p > .05$ for all comparisons); instead we observed a significant drop in heart rate with respect to baseline in both groups in the time window ranging from 1.88 to 4.45 seconds after the chatbot's response onset (Welch *t*-value ranged from $t = -2.05$, $p = .0499$ to $t = -3.48$, $p = .0016$; see Fig. 9). This may indicate that participants were emotionally and physiologically preparing for the chatbot's answer, then realised that the answer was not threatening or irritating enough to trigger a stress reaction.

4.2.3. EMG results

The between-group comparison of the EMG signal evoked by the chatbot response yielded significant differences only for the frown muscle (*corrugator supercilii*). The amplitude of the corrugator's EMG signal was significantly higher for the AVATAR than the TEXT group in two time windows. The first time window started at 1.84 and lasted up to 2.32 seconds after the chatbot's response onset (*t* values ranged from $t = -2.06$, $p = .0489$ to $t = -2.46$, $p = .0205$). The

second time window started at 3.23 and lasted up to 5.13 seconds after the chatbot's response onset (*t* values ranged from *t* = -2.06, *p* = .0486 to *t* = -3.07, *p* = .0048; see Fig. 10).

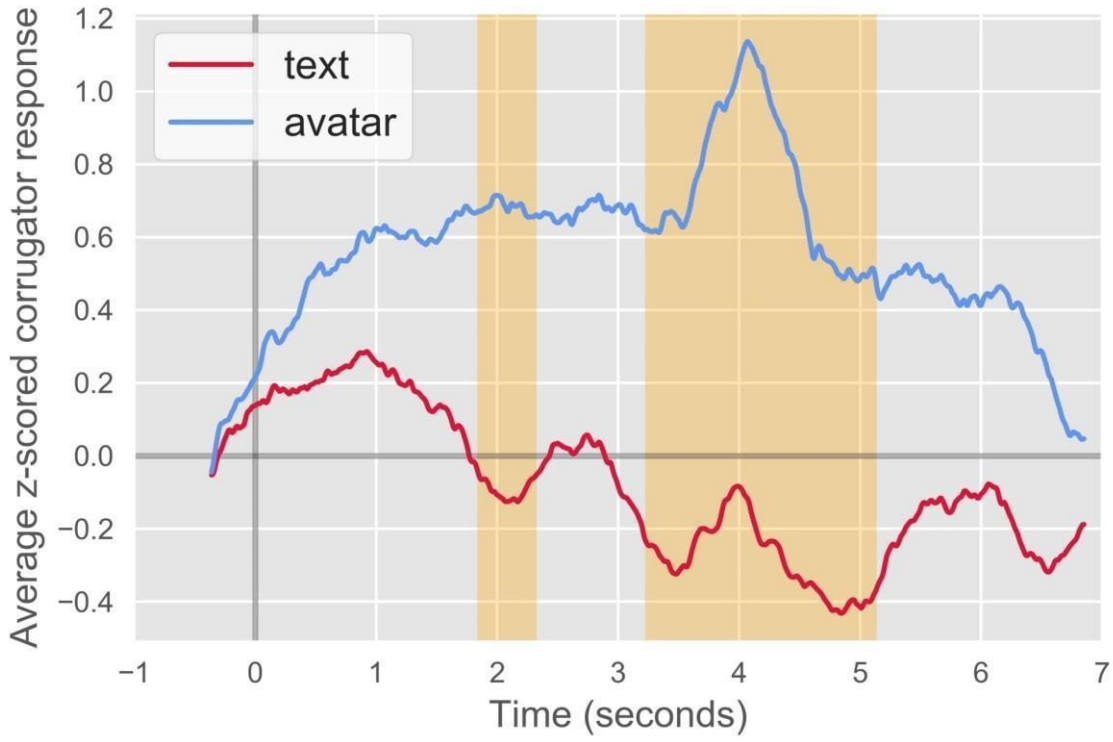

Figure 10. Time-resolved changes in *corrugator supercilii* muscle electrical activity for both experimental groups. Significant between-group differences are marked with orange.

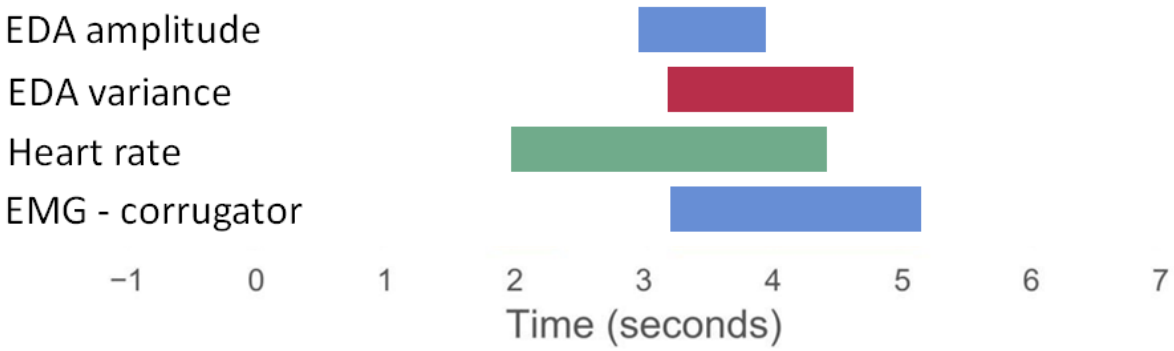

Figure 11. Different psychophysiological measures presented on one chart. Colours indicate which group had higher results in the corresponding measure: red—TEXT; blue—AVATAR; and green—no difference between groups, but both groups differed significantly from zero in their evoked response.

4.2.4. Psychophysiological measures and questionnaires

We also investigated correlations between various psychophysiological indices and questionnaire results in order to determine the possible relationship between these measures.

Figure 12 presents correlations between the EDA and competence factor for both groups (there were no significant differences between the groups) . The correlation is significant in the two time windows: first, ranging from -0.29 to 0.55 seconds after the chatbot's response onset (Pearson correlation coefficient ranged from r = -.37, p = .0498 to r = -.47, p = .0104); second, ranging from 2.12 to 3.56 seconds (Pearson correlation coefficient ranged from r = -.37, p = .0499 to r = -.47, p = .0104). This result indicates that the less competent a chatbot seemed, the higher electrodermal response it caused, thereby triggering a more intense emotional reaction.

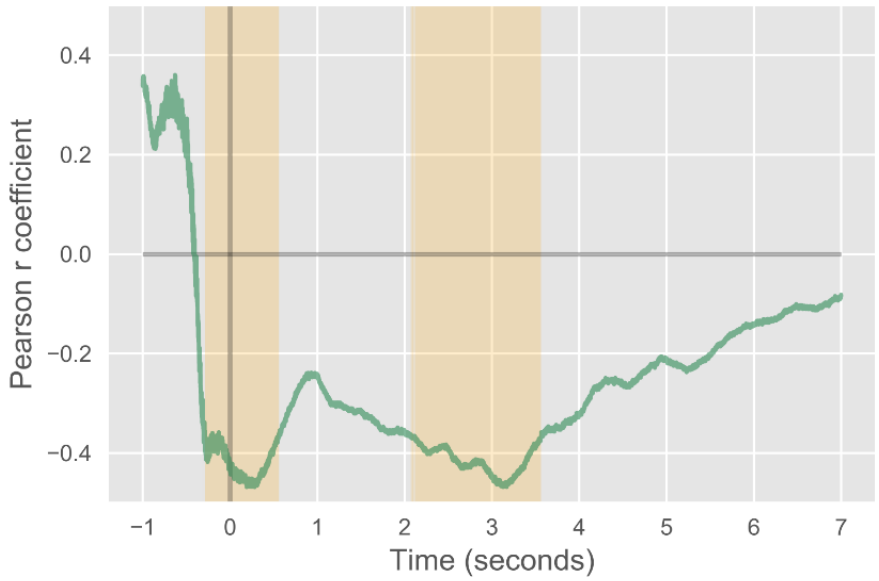

Figure 12. Correlation between EDA and the competence factor for both groups. Yellow indicates time windows where the correlation is significant.

Next, we correlated the EDA signal with the uncanny valley factor, uncovering a strong significant correlation for the AVATAR group in the late time window starting at 5.55 seconds after the chatbot's response onset and lasting at least until the segment's end (7.00 seconds after the chatbot's response onset), where minimum correlation coefficient was r = .55, p = .0498, and maximum correlation coefficient was r = .70, p = .0071 (see Fig. 13). The result indicates that, in the AVATAR group, participants with higher EDA amplitude in this time window rated the chatbot higher on the uncanny valley factor scale (which means they considered the chatbot to be more "weird").

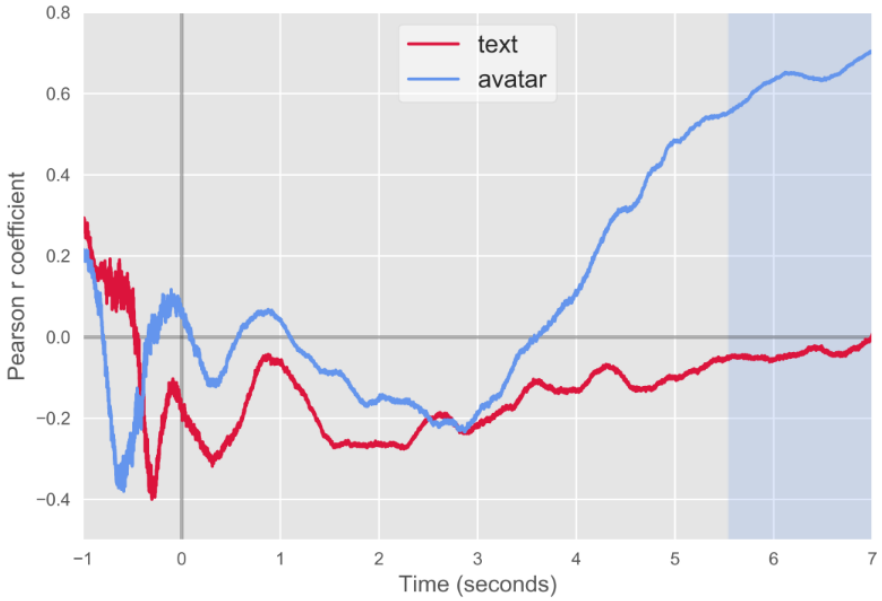

Figure 13. Correlation between EDA and the uncanny valley factor with EDA. Blue indicates the time window where the correlation coefficient for the AVATAR group is significantly different from zero.

5. Discussion

Our overall assessment of the psychophysiological signals and the questionnaires is that participants enjoyed their interaction with both types of chatbots, with the reservation that they viewed the TEXT chatbot more positively. Moreover, participants expressed an expectation of more frequent interactions of that kind in the future, stressing the need to develop bots' communication abilities. The most significant differences between the groups occurred, as we expected, in the uncanny valley factor. The interaction with the TEXT chatbot was more pleasant for participants, compared to the interaction with the AVATAR chatbot. These two factors were strongly correlated, as we showed in the analysis section; in accordance with previous research, this naturally indicates that the "weirder" the bot is perceived to be, the more intense the negative affect it will cause. Perhaps the slightly unnatural voice and animation of the avatar were viewed as elements unsuccessfully imitating a human. Moreover, greater expectations related with the multi-channel experience resulted in greater dissatisfaction with the AVATAR bot's performance. These results suggest that chatbots should not be designed to pretend to be human—at least not if they do so ineptly.

EDA results (Fig. 6) showed a difference between both groups just before and around the onset of the chatbot's response with the AVATAR group, achieving a higher rate of physiological arousal. This may indicate participants' expectation of the emotional or cognitive content of a chatbot's upcoming response. Participants from the AVATAR group may expect a more emotional content of the chatbot's response. The significant between-group difference in the EDA variance, with it being higher in the TEXT group, may indicate that the emotional arousal in this group depended not on the audio-visual features of the chatbot, but rather on the message contents. The chatbot group manifested a higher amplitude of the EDA and a lower

variance; hence, we suppose that the audio-visual features of the chatbot "channelled" and directed the emotional arousal.

In terms of the heart rate, we observed general differences between the groups, with HR being significantly higher in the AVATAR group (Fig. 8). This result may suggest that participants interacting with the TEXT chatbot were less emotionally aroused during the chat session. At the same time, HR dropped in both groups during the interactions in the time window from the second to the fourth second in reference to the chatbot's response onset. Previous studies have shown that abrupt HR decreases can co-occur together with increases in sympathetic system activity [44–46].

EMG activity showed significant differences in the EMG amplitude recorded from the frown muscle, with stronger amplitude in the AVATAR group. This result may indicate that, in this group, participants expressed more negative affect than their counterparts from the TEXT group [47,48]. However, this result may also suggest that participants in this group experienced an implicit need for affiliation with the AVATAR chatbot [49].

Interesting effects can also be observed in the correlations between psychophysiological measures and questionnaire factors (competence and uncanny valley). The competence factor correlated negatively with EDA for both groups in the time window from the 2 second to the 3.5 second (Fig. 12), which means that participants were more physiologically aroused when the chatbot was assessed as more incompetent. At the same time, the uncanny valley factor correlated positively with EDA only in the AVATAR group in the time window between seconds 5.5 and 7, which means that the more uncanny or "weird" the AVATAR chatbot seemed for the participant, the more physiologically agitated the participant was.

These results support our hypothesis stating that people are more physiologically aroused when they encounter a being imperfectly imitating a human (as in the case of the AVATAR chatbot), which concurs with findings from similar studies [50],

The psychophysiological data indicated systematic differences between groups, determined by the type of the chatbot. The AVATAR chatbot triggered more intense emotional reactions than the TEXT one. The EDA, HR, and EMG results support this claim. Most of these between-group psychophysiological differences occurred within the time window ranging from the third to fifth second after the chatbot's response onset. This timing agreement between psychophysiological measures (see Fig. 11) may suggest a common underlying psychological process that manifests itself in EDA, HR, and EMG, like increase in the sympathetic system activity. It is also possible that this psychological process begins earlier than the psychophysiological signal indicates (due to natural delay of it, as discussed in the literature [40]). Therefore, further studies using electroencephalography (EEG) might be useful. Previous studies have demonstrated that it is possible to track emotional valence and arousal, with the usage of specific EEG indices, like late positive potential [50].

6. Limitations

Although our research reached its aims, there were some unavoidable limitations and shortcomings. In our experiment, we used a combination of overlapping text, sound, and video (avatar) for the experimental group. However, in the next round of experiments, we will make an attempt to eliminate overlapping channels of communication in order to get a clearer understanding of the medium and the extent to which it possibly elicits negative affect. Other limitations include a lack of distinction between the chatbot's voice and visual representation in order for users to get used to it instead of being surprised. In other words, the psychophysiological and/or behavioural differences between the groups could have been produced by one of the auditory channels (the "avatar chatbot" differed from the "simple chatbot" by including a visual representation; it also presented its responses not only on screen in text form, but also by reading them aloud).

We abstained from including a human interlocutor in the experiment scheme because the experimental control of such a group would be even more difficult than controlling conversations between a chatbot and a human, which is in itself already stochastic in nature (there is an uncountable variability of possible conversation scenarios between these subjects). In the literature on human–computer or human–robot interactions, using human subjects as one of the experiment conditions or groups is rare [32] and is burdened with relatively uncontrollable variability. In addition, using, for example, a human shadower of a chatbot in an experiment paradigm causes delays in communication and feels artificial [9].

In future research, it is also advisable to design a methodology that takes into account the expectations of users and analyse their correlations with multichannel bot experience. Future research should also include the usage of devices that do not hinder or distract from the natural flow of communication. If portable neuroimaging and psychophysical devices turn out to have adequate accuracy, replacing electrodes with sensors that do not require touch, as well as using techniques related to camera-supported facial expressions recognition, may become a viable methodology. Such experiments could also be conducted outside the lab and thus have the potential to be more similar to everyday interactions with chatbots that users encounter.

A question that could and should be resolved soon is the issue of emotion expression through voice in bots. An interesting study performed by scholars from MIT Social Robotics Lab showed the important role voice plays in real-time interactions between humans (in this case children) and non-humans [51]. The shallowness of voice and lack of emotional expression are crucial factors in assessing attitudes towards a machine. Therefore, further qualitative, psychophysiological, and electrophysiological studies should follow. To date, most bots have expressed very rare emotions through the audio channel. However, with the development of speech synthesis, this may change, potentially affecting the psychophysiological responses to bots in general and the uncanny valley effect in particular.

7. Prospective research

There is a need to gradually test the increasingly sophisticated bots and social robots using various methodologies, including psychophysiological ones. An important approach of future research would be to include the addition of another experimental group with a human–human interaction, where the chatbot is replaced by either a recorded or live human being interacting with study participants. Another point worth considering would be to allow future experiment participants to choose which chatbot they would like to interact with and further assess the uncanny valley effect while taking their initial choice into consideration. As we already know from the body of research on human–machine interaction, users tend to differ significantly in terms of preferred channels and style of communication [52] [53]. We decided not to embed a human–human control group in the experiment as our extensive literature review showed that such groups would be rather incomparable at this stage of technological readiness in the context of bots' realism. Research thus far has demonstrated that conversation styles and duration between human–human and human–chatbot interactions differed very significantly [54].

At this point, our project is intended to scale up from the simple bots like the one designed for the experiment to more refined commercial and non-commercial bots based on deep learning, which again use multi-channel communication methods. However, at the same time, our research team has started the pilot process of developing an experimental protocol of comparing chatbots with human help-desk workers, which is a necessary—albeit very challenging—next step in order to reliably research the attitude of the general public towards chatbots. As bots become increasingly more popular in the professional and personal sphere, the task of understanding how they are perceived and what drives this perception becomes urgent and necessary.

Acknowledgements

Dr Aleksandra Przegalinska was supported by the Polish Ministry of Science grant (no. DN/MOB/102/IV/2015).

No conflict of interest needs to be declared.

All authors contributed to the article. L. Ciechanowski and M. Magnuski prepared and conducted the experiment and data analyses. A. Przegalinska and P. Gloor worked out the theoretical part of the study and the article as well as the experiment's general concept.

We would like to express our gratitude towards the InteliWISE company owners for providing the chatbots used in our experiment.

We are also grateful for the insightful comments and remarks by the two anonymous reviewers.

References

[1] R.R. Hightower, L.T. Ring, J.I. Helfman, B.B. Bederson, J.D. Hollan, Graphical Multiscale Web Histories: A Study of Padprints, in: Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia : Links, Objects, Time and Space---Structure in Hypermedia Systems: Links, Objects, Time and Space---Structure in Hypermedia Systems, ACM, New York, NY, USA, 1998: pp. 58–65.

[2] E. Hutchins, Cognition in the Wild, MIT Press, 1995.

[3] E.I. Barakova, Social Interaction in Robotic Agents Emulating the Mirror Neuron Function, in: Nature Inspired Problem-Solving Methods in Knowledge Engineering, Springer, Berlin, Heidelberg, 2007: pp.

389–398.
[4] M.-C. Jenkins, R. Churchill, S. Cox, D. Smith, Analysis of User Interaction with Service Oriented Chatbot Systems, in: Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, Springer, Berlin, Heidelberg, 2007: pp. 76–83.
[5] B. Reeves, C. Nass, How people treat computers, television, and new media like real people and places, CSLI Publications and Cambridge. (1996). http://www.humanityonline.com/docs/the%20media%20equation.pdf.
[6] K. Yun, K. Watanabe, S. Shimojo, Interpersonal body and neural synchronization as a marker of implicit social interaction, Sci. Rep. 2 (2012) 959.
[7] J. Decety, C. Lamm, The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition, Neuroscientist. 13 (2007) 580–593.
[8] K. Sung, S. Dolcos, S. Flor-Henry, C. Zhou, C. Gasior, J. Argo, F. Dolcos, Brain imaging investigation of the neural correlates of observing virtual social interactions, J. Vis. Exp. (2011) e2379.
[9] K. Corti, A. Gillespie, A truly human interface: interacting face-to-face with someone whose words are determined by a computer program, Front. Psychol. 6 (2015) 634.
[10] G. Hofree, P. Ruvolo, M.S. Bartlett, P. Winkielman, Bridging the mechanical and the human mind: spontaneous mimicry of a physically present android, PLoS One. 9 (2014) e99934.
[11] J. Kacprzyk, S. Zadrozny, Computing With Words Is an Implementable Paradigm: Fuzzy Queries, Linguistic Data Summaries, and Natural-Language Generation, IEEE Trans. Fuzzy Syst. 18 (2010) 461–472.
[12] K. Morrissey, J. Kirakowski, "Realness" in Chatbots: Establishing Quantifiable Criteria, in: International Conference on Human-Computer Interaction, Springer, 2013: pp. 87–96.
[13] J. Weizenbaum, ELIZA—a computer program for the study of natural language communication between man and machine, Commun. ACM. 9 (1966) 36–45.
[14] J. Weizenbaum, J. McCarthy, Computer power and human reason: From judgment to calculation, (1977).
[15] R. Wilensky, Planning and understanding: A computational approach to human reasoning, (1983). http://www.osti.gov/scitech/biblio/5673187 (accessed June 5, 2017).
[16] V.R. Basili, R.W. Selby, D.H. Hutchens, Experimentation in software engineering, IEEE Trans. Software Eng. SE-12 (1986) 733–743.
[17] B. Batacharia, D. Levy, R. Catizone, A. Krotov, Y. Wilks, CONVERSE: a Conversational Companion, in: Y. Wilks (Ed.), Machine Conversations, Springer US, 1999: pp. 205–215.
[18] B.A. Shawar, E. Atwell, Using dialogue corpora to train a chatbot, in: Proceedings of the Corpus Linguistics 2003 Conference, 2003: pp. 681–690.
[19] H. Mark, Battle of the digital assistants: Cortana, Siri, and Google Now, PC World. 13 (2014).
[20] E. Moemeka, E. Moemeka, Leveraging Cortana and Speech, in: Real World Windows 10 Development, Apress, 2015: pp. 471–520.
[21] A. Hayes, Amazon Alexa: A Quick-start Beginner's Guide, CreateSpace Independent Publishing Platform, 2017.
[22] P. Rane, V. Mhatre, L. Kurup, Study of a home robot: Jibo, in: International Journal of Engineering Research and Technology, IJERT, 2014.
[23] E. Guizzo, Cynthia Breazeal Unveils Jibo, a social robot for the home, IEEE Spectrum. (2014).
[24] K.F. MacDorman, T. Minato, M. Shimada, Assessing human likeness by eye contact in an android testbed, Proceedings of the. (2005). http://www.psy.herts.ac.uk/pub/SJCowley/docs/humanlikeness.pdf.
[25] M. Mori, Bukimi no tani [the uncanny valley], Energy. 7 (1970) 33–35.
[26] M.L. Walters, D.S. Syrdal, K. Dautenhahn, R. te Boekhorst, K.L. Koay, Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion, Auton. Robots. 24 (2008) 159–178.
[27] J. 'ichiro Seyama, R.S. Nagayama, The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces, Presence: Teleoperators and Virtual Environments. 16 (2007) 337–351.
[28] J.R. Shaffer, E. Orlova, M.K. Lee, E.J. Leslie, Z.D. Raffensperger, C.L. Heike, M.L. Cunningham, J.T. Hecht, C.H. Kau, N.L. Nidey, L.M. Moreno, G.L. Wehby, J.C. Murray, C.A. Laurie, C.C. Laurie, J. Cole,

T. Ferrara, S. Santorico, O. Klein, W. Mio, E. Feingold, B. Hallgrimsson, R.A. Spritz, M.L. Marazita, S.M. Weinberg, Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology, PLoS Genet. 12 (2016) e1006149.
[29] K.F. MacDorman, H. Ishiguro, The uncanny advantage of using androids in cognitive and social science research, Interact. Stud. 7 (2006) 297–337.
[30] K.F. MacDorman, Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it, in: CogSci-2005 Workshop: Toward Social Mechanisms of Android Science, 2005: pp. 106–118.
[31] D. Hanson, Exploring the aesthetic range for humanoid robots, in: Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, Citeseer, 2006: pp. 39–42.
[32] J.-D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P.F. Dominey, J. Ventre-Dominey, I Reach Faster When I See You Look: Gaze Effects in Human–Human and Human–Robot Face-to-Face Cooperation, Front. Neurorobot. 6 (2012). doi:10.3389/fnbot.2012.00003.
[33] A. Gillespie, K. Corti, The Body That Speaks: Recombining Bodies and Speech Sources in Unscripted Face-to-Face Communication, Front. Psychol. 7 (2016) 1300.
[34] F. Schrammel, S. Pannasch, S.-T. Graupner, A. Mojzisch, B.M. Velichkovsky, Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience, Psychophysiology. 46 (2009) 922–931.
[35] A.J. Fridlund, J.T. Cacioppo, Guidelines for human electromyographic research, Psychophysiology. 23 (1986) 567–589.
[36] B.M. Appelhans, L.J. Luecken, Heart rate variability as an index of regulated emotional responding, Rev. Gen. Psychol. 10 (2006) 229.
[37] G. Pochwatko, J.-C. Giger, M. Różańska-Walczuk, J. Świdrak, K. Kukiełka, J. Możaryn, N. Piçarra, Polish Version of the Negative Attitude Toward Robots Scale (NARS-PL), Journal of Automation Mobile Robotics and Intelligent Systems. 9 (2015). https://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-70bedbb8-141c-40e6-b3b9-596032d8abda/c/Pochwatko_polish_version.pdf.
[38] T. Fong, I. Nourbakhsh, K. Dautenhahn, A Survey Of Socially Interactive Robots: Concepts, Design And Applications, Technical Report No. Cmu—Ri—Tr—02—29, Carnegie Mellon University, 2002. https://pdfs.semanticscholar.org/a764/15c475a8e40ded6697982ebbe7b6141505ca.pdf.
[39] S.T. Fiske, A.J.C. Cuddy, P. Glick, J. Xu, A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition, J. Pers. Soc. Psychol. 82 (2002) 878–902.
[40] D.R. Bach, G. Flandin, K.J. Friston, R.J. Dolan, Modelling event-related skin conductance responses, Int. J. Psychophysiol. 75 (2010) 349–356.
[41] M. Delacre, D. Lakens, C. Leys, Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test, International Review of Social Psychology. 30 (2017). http://rips.ubiquitypress.com/articles/10.5334/irsp.82/.
[42] C. Carreiras, A.P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred, BioSPPy - Biosignal Processing in Python, 2015. https://github.com/PIA-Group/BioSPPy (accessed January 3, 2018).
[43] P. Hamilton, Open source ECG analysis, in: Computers in Cardiology, 2002: pp. 101–104.
[44] I.B. Mauss, M.D. Robinson, Measures of emotion: A review, Cogn. Emot. 23 (2009) 209–237.
[45] M.M. Bradley, P.J. Lang, Measuring emotion: Behavior, feeling, and physiology, Cognitive Neuroscience of Emotion. (2000). https://books.google.com/books?hl=en&lr=&id=A2s963AzymYC&oi=fnd&pg=PA242&dq=Measuring+emotion+Behavior+feeling+physiology+Bradley+Lang&ots=m9Q4e83Bs7&sig=3aftBm6ktk0vuYJB9XjMUmJphgk.
[46] P.J. Lang, M.M. Bradley, B.N. Cuthbert, Motivated attention: Affect, activation, and action, in: P.J. Lang (Ed.), Attention and Orienting: Sensory and Motivational Processes, books.google.com, 1997.
[47] J.T. Cacioppo, R.E. Petty, M.E. Losch, H.S. Kim, Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions, J. Pers. Soc. Psychol. 50 (1986) 260–268.

[48] S. Topolinski, F. Strack, Corrugator activity confirms immediate negative affect in surprise, Front. Psychol. 6 (2015) 134.
[49] A. Kordik, K. Eska, O.C. Schultheiss, Implicit need for affiliation is associated with increased corrugator activity in a non-positive, but not in a positive social interaction, J. Res. Pers. 46 (2012) 604–608.
[50] M. Cheetham, L. Wu, P. Pauli, L. Jancke, Arousal, valence, and the uncanny valley: psychophysiological and self-report findings, Front. Psychol. 6 (2015) 981.
[51] S. Druga, R. Williams, C. Breazeal, M. Resnick, "Hey Google is it OK if I eat you?," in: Proceedings of the 2017 Conference on Interaction Design and Children - IDC '17, 2017. doi:10.1145/3078072.3084330.
[52] M. Xuetao, F. Bouchet, J.-P. Sansonnet, Impact of agent's answers variability on its believability and human-likeness and consequent chatbot improvements, in: Proc. of AISB, 2009: pp. 31–36.
[53] S. A., D. John, Survey on Chatbot Design Techniques in Speech Conversation Systems, Ijacsa. 6 (2015). doi:10.14569/IJACSA.2015.060712.
[54] J. Hill, W. Randolph Ford, I.G. Farreras, Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations, Comput. Human Behav. 49 (2015) 245–250.