User Perceptions of Extraversion in Chatbots after Repeated Use

Sarah Theres Völkel
LMU Munich
Munich, Germany
sarah.voelkel@ifi.lmu.de

Ramona Schoedel
LMU Munich
Munich, Germany
ramona.schoedel@psy.lmu.de

Lale Kaya
LMU Munich
Munich, Germany
l.kaya@campus.lmu.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

ABSTRACT

Whilst imbuing robots and voice assistants with personality has been found to positively impact user experience, little is known about user perceptions of personality in purely text-based chatbots. In a within-subjects study, we asked N=34 participants to interact with three chatbots with different levels of Extraversion (extraverted, average, introverted), each over the course of four days. We systematically varied the chatbots' responses to manipulate Extraversion based on work in the psycholinguistics of human behaviour. Our results show that participants perceived the extraverted and average chatbots as such, whereas verbal cues transferred from human behaviour were insufficient to create an introverted chatbot. Whilst most participants preferred interacting with the extraverted chatbot, participants engaged significantly more with the introverted chatbot as indicated by the users' average number of written words. We discuss implications for researchers and practitioners on how to design chatbot personalities that can adapt to user preferences.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Natural language interfaces**.

KEYWORDS

chatbot, conversational agent, extraversion, personalisation, personality

ACM Reference Format:
Sarah Theres Völkel, Ramona Schoedel, Lale Kaya, and Sven Mayer. 2022. User Perceptions of Extraversion in Chatbots after Repeated Use. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3491102.3502058

1 INTRODUCTION

When we meet another person, we immediately and automatically form an impression of their personality, which significantly influences our future behaviour and expectations towards that person [75]. Likewise, conversational agents (CAs), such as voice assistants and chatbots, are perceived as social actors and, therefore, elicit similar personality judgements [81, 84, 99]. That is, users subconsciously assign CAs a personality regardless of whether this was intended by the conversation designer or not. Similar to human-human interaction, this personality assignment has been shown to influence user trust [19, 133], affection [13, 14, 19, 82, 125], engagement [112, 133], self-disclosure [45, 133], and purchase behaviour [112]. However, despite their ubiquity, commercially available CAs have so far taken a one-size-fits-all approach, ignoring the potential benefits that deliberate personality manipulation and personalisation may bring.

A reason for this could be that systematically imbuing CAs with personality is challenging [98]. For example, the Google Assistant developer guide suggests listing four to six key adjectives that describe the CA's personality and then design dialogues based on characters that reflect those adjectives[1]. This process does not give the developer any insights as to how the CA should behave to convey the intended personality adjectives. To address this challenge, previous efforts in research on how to manipulate robot and voice assistant personality have drawn from a plethora of work in psycholinguistics on the relationship between personality and non-verbal human behaviour cues, such as speech rate, gaze, and proximity (e.g. [8, 23, 82]). Whilst robots and voice assistants can leverage this myriad of non-verbal behaviour cues, little is known about whether verbal cues alone are sufficient to evoke the intended personality traits. Moreover, personality perceptions of CAs have been predominantly evaluated for short one-time interactions only. However, examining the attribution of personality after repeated use is particularly crucial as personality perceptions can change after prolonged contact [9] and personality adaptation of CAs seems primarily meaningful in the context of long-term interaction [102].

We embark on this research gap by presenting a systematic manipulation and evaluation of personality-imbued chatbots in the context of an app for stress tracking after repeated use. Because the relationship between perceptible behaviour cues and the personality trait Extraversion is most pronounced [94], we manipulated three different levels of extraverted chatbots, introverted, average, and extraverted, by systematically varying the chatbots' use of language. In a within-subjects study, we asked N=34 participants to converse with our three Telegram chatbots, each over the course of four days, in order to examine users' perception of the chatbot personalities after repeated, prolonged use. We analysed participants' subjective preferences and the number of words exchanged with the chatbots. Finally, we examined the relationship between the user's personality and their chatbot personality preference. In summary, we address the following three research questions:

RQ1 *Can different levels of Extraversion be synthesised for solely text-based chatbots using verbal cues from human behaviour?*

[1]https://developers.google.com/assistant/conversation-design/create-a-persona, last accessed April 10, 2022

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3502058

RQ2 *Which level of Extraversion do users prefer after repeated use of the chatbots?*

RQ3 *Does user personality influence user preference for the chatbot personality?*

Our contribution is twofold: First, we present a set of verbal cues derived from psycholinguistic literature to induce different levels of Extraversion implemented in a chatbot app. Second, we present empirical findings as to how the chatbots' personalities influence user preference and interaction behaviour. Our results show that participants perceived the extraverted and average chatbots as such, whereas verbal cues transferred from human behaviour were insufficient to create a chatbot perceived as introverted. Whilst most participants preferred interacting with the extraverted chatbot, participants engaged significantly more with the introverted chatbot, as indicated by the users' average number of written words. These findings provide much needed information to researchers and practitioners on how to design chatbot personalities that can adapt to the user's preference.

2 RELATED WORK

Below we summarise works on human personality and their markers in language, how to imbue conversational agents with personality, how to adapt the conversational agent personality to the user, and conversational agents in the context of mental health care.

2.1 Human Personality

Human personality is defined by individual, consistent, and lasting patterns of behaviour, cognition, and emotions [77]. To formally describe personality, the *Big Five* model, also termed *OCEAN* or *Five Factor* theory, has emerged as the most prevalent paradigm in scientific research [74]. The Big Five model comprises five broad dimensions, which are composed of several sub-facets [30, 46, 75]:

Openness reflects a tendency to be imaginative, artistically interested, creative, intellectually curious, and open-minded. *Conscientiousness* reflects a tendency to be disciplined, orderly, dutiful, competent, ambitious, and cautious. *Extraversion* reflects a tendency to be friendly, sociable, assertive, dynamic, adventurous, and cheerful. *Agreeableness* reflects a tendency to be trustful, genuine, helpful, modest, obliging, and cooperative. *Neuroticism* reflects a tendency to be anxious, easily stressed, depressed, and emotionally unstable.

Previous research has demonstrated that these traits are relatively stable across different situations and cultures [35, 75], and change only moderately over a lifespan [26, 29]. The Big Five have also been leveraged in Human-Computer Interaction (HCI) research to describe differences in how CAs and robots express behaviour [81, 108, 117, 119, 124].

2.2 Personality Markers in Language

People externalise their personality through observable behaviour. As personality is a latent and abstract construct, individuals can only infer a person's personality traits from these perceptible behaviour cues [107, 120]. Behaviour cues that individuals consistently associate with a particular personality disposition are termed *personality markers* [106]. These personality markers can be verbal (*what* is said, e.g. speech or written text) or non-verbal (*how* it is said, e.g. gestures, speech rate) [9]. In this paper, we focus on text-based chatbots for which only the verbal output can be manipulated.

The relationship between human personality and language use has a longstanding tradition in psychology and linguistics, demonstrating that language use greatly varies from person to person but is internally consistent and reliable [17, 24, 25, 38, 43, 80, 88, 93, 94, 105, 106]. To identify personality markers, the relationship between individuals' verbal output and their self-reported personality via standardised questionnaires was usually analysed by means of correlations [54, 94] or group comparisons [44]. For example, verbal personality markers were found for word use [44, 54, 94], phrases [44], and linguistic style [12, 43, 88] by examining written essays [44, 94], self-narratives [54], and text messages [55].

Most personality markers were found for the personality dimension Extraversion [106]. Extraverted people are characterised by a higher total verbal output; that is, they talk and write more than introverts [24, 25, 80, 93, 105]. Furthermore, people high in Extraversion tend to use an implicit and abstract speech style with a simple sentence structure, falling back upon a sparse vocabulary with highly frequent words [12, 38, 43, 88]. Conversely, introverted individuals tend to speak more explicitly, formally, and concretely, expressing themselves with complicated sentences and a rich vocabulary of infrequent words [12, 38, 43, 88]. People's level of Extraversion is also reflected in their choice of words [44, 54, 94]. Finally, extraverted individuals are more prone to use emojis containing hearts [123] or signifying happiness [73]. In this paper, we will leverage these personality markers to imbue chatbots with three different levels of Extraversion.

2.3 Imbuing CAs with Personality

HCI researchers have imitated the aforementioned behaviour cues associated with human personality to deliberately manipulate CA personality [41]. So far, the effect of systematic behaviour variations on personality impressions has mainly been researched for robots [6, 8, 20, 102], virtual embodied agents [3, 7, 23, 27, 57, 61, 79, 128], and speech-based CAs such as voice assistants [19, 81, 83, 85, 89, 96]. These evaluations focused on manipulating non-verbal cues, such as gaze [8], gestures [20, 61], posture [57], proximity [23], and paraverbal cues, such as speech rate and pitch [82]. Conversely, little attention was paid to whether personality in chatbots can manifest itself through verbal cues alone.

Zhou et al. [133] created different chatbots by variations in verbal cues and evaluated them through simple keywords given by participants. Shumanov and Johnson [112] successfully manipulated a chatbot's choice of words to design an introverted and extraverted customer support chatbot. They performed a dichotomous manipulation check by having experts and users classify the chatbots' responses as either extra- or introverted. Ruane et al. [103] also manipulated the choice of words along with complementary verbosity, thereby creating a chatbot that is both high in Extraversion and Agreeableness, and one that is low in these two dimensions. To evaluate the personality perception, they asked participants to openly describe the two chatbots' personalities.

Previous works compared two opposing versions of Extraversion; however, it has not been evaluated how strongly extraverted they are perceived to be on a continuous scale. This evaluation seems particularly important because prior work on robots [8] and voice interfaces [82] has shown that users prefer different levels of Extraversion in these agents based on their own personality. As the distribution of human personality traits is expected to follow a Gaussian curve in the population [78], we assume that an *average* shade of Extraversion in chatbots might be beneficial to the user.

In this paper, we examine if three different levels of Extraversion can be created by adapting verbal cues in the chatbots' responses. To this end, we not only collect open personality descriptions, but also evaluate the targeted manipulation by means of a standard personality questionnaire. Moreover, our work explores if these three levels of personality are indeed beneficial by examining user preferences for all three levels.

2.4 Adapting CAs to the User

Users' personality perceptions determine their attitude and behaviour towards the CA [81]. Previous work has noted individual differences in user preferences for CA personality [19, 65, 82, 122, 125], with users tending to favour personalities that match their own [8, 82, 116], termed the *similarity attraction effect* [22, 82].

For chatbots, Völkel and Kaya [124] demonstrated that users high in Agreeableness preferred a highly agreeable chatbot but they did not detect a reversed relationship for low Agreeableness. Shumanov and Johnson [112] found out that a congruent level of Extraversion in user and chatbot personality had a positive impact on user engagement with the commerce chatbot and purchasing outcomes for interactions involving social gain. Echoing these findings, Gnewuch et al. [45] showed that corresponding dominance levels in chatbot and user increased user self-disclosure of personal information during the conversation. Research on robots and speech-based CAs corroborates these findings, in particular for the personality dimension Extraversion [8, 14, 82]. For example, extraverted users showed a preference for an extraverted voice user interface on a book buying website [82] as well as a virtual real estate agent engaging in social talk [14].

However, the effect of CA personalities on user preferences and behaviour has so far only been examined for one-time, short interactions, although user attitude will likely change with repeated use [125]. Hence, in this work, we examine user engagement, gauged by the number of words written by participants, and preference after four days of interacting with the chatbot.

2.5 CAs for Mental Health Support

We situated our personality-imbued chatbots in the context of a stress tracking and reflection application due to the growing availability of CAs for health-related support [59, 72]. Using CAs for (mental) health support has circulated as an idea since *ELIZA*, the very first psychotherapist chatbot [129], and recent enhancements in CA capabilities have sparked a new wave of research in the HCI community [11, 28, 68]. Despite severe consequences [91], more than 80% of people suffering from mental health conditions are without treatment [91] and lack qualified mental health practitioners, creating a bottleneck in mental health care [11, 71]. Lately, the COVID-19 pandemic has aggravated the situation as the need for mental health service has increased rapidly and unexpectedly whilst therapist face-to-face meetings were often impossible [72, 90, 92]. In light of these shortcomings, CAs have the potential to open up new frontiers in mental health treatment [59, 64, 72, 115] as they offer users help regardless of time and location, at low cost, and with less fear of being judged [15, 64, 69, 71]. Despite these advantages, current "one-size-fits-all" mental health applications often fail, underlining the need for systems that are tailored to individual user needs to increase pertinence, user engagement, treatment adherence, and clinical efficacy [42, 50, 52, 62, 104, 110, 132]. However, personalisation of CAs in health care is still rare and often lacks a theoretical framework [60]. Hence, infusing CAs for health support with personality has been highlighted as an opportunity to improve user engagement [2] but to the best of our knowledge the relationship between user personality and preference for CA personality has not been investigated.

3 CHATBOT DESIGN

We designed three chatbots that exhibit different levels of Extraversion (introverted, average, and extraverted) through targeted manipulation of their language. We situated the chatbots in a stress tracking and reflection application and asked participants to interact with the three chatbots for four days each (three days of tracking stress and the fourth day for feedback and reflection). In this section, we first describe the conversation flow between chatbot and user. We then point out how this conversation flow is realised for the three different chatbot versions to manipulate different personality perceptions. Finally, we briefly explain the implementation of the chatbots as a Telegram messaging bot. To ensure the chatbots' functionality and comprehensibility, we pilot-tested the chatbots in several iterations.

3.1 Conversation Flow

To ensure that the three chatbots maintained their predefined personality, we followed a guided conversation approach (in contrast to an open conversation), with the respective questions and responses being purely script-based. Upon start, each chatbot welcomes the user, introduces itself, and explains the study procedure. Afterwards, the chatbot prompts the user to answer questions about their perceived stress following the Daily Inventory of Stressful Events (DISE) questionnaire [4], which we describe in the next subsection. On the following interaction days, we adopted an experience sampling approach, with the chatbot actively triggering the user daily via a push-notification to fill out the DISE questionnaire at a random point within a user-set time frame. If the user does not answer within two hours, then the chatbot sends a reminder. Thereby, we collect in situ self-reports of user stressors [118].

3.1.1 Daily Inventory of Stressful Events. The DISE was originally composed as a semi-structured interview instrument that combines participants' subjective perception of the severity of a stressor and more objective criteria such as the type of threat (e.g. danger or disappointment) [4]. The inventory comprises seven stressor questions that ascertain the occurrence of stressful events across different life domains via *yes/no* answers (e.g. "Did you have an argument or disagreement with anyone since this time yesterday?"). For each affirmative stressor question, participants are probed with several follow-up questions:

(1) A series of open-ended probe questions that collect the participant's description of the stressful event (e.g. "What happened and what about it would most people consider stressful?")
(2) A Likert scale question pertaining the perceived severity of the stressful event ("How stressful was this for you?").
(3) If the participant perceives the event as stressful, then they are presented with a list of seven Likert scale primary appraisal questions as to which of their goals and values were at risk due to the stressor (e.g. "How much did it risk your physical health or safety?").

3.1.2 Feedback Report. On the fourth interaction day, the user does not answer any questions but the chatbot sends them a feedback report detailing the user's perceived stressors and their severity as visualised by a bar and line chart, respectively. Furthermore, the chatbot performs a sentiment analysis on users' responses to the open-ended probe questions, illustrated with another line chart. For each chart, the chatbot includes a brief explanation. If the user did not experience a stressful event on any of the days, then the chatbot points this out and discloses how it arrived at this conclusion. Following the report, the chatbot requests the user to reflect on their perceived stress and asks for feedback on the report's accuracy.

3.1.3 Deviations from Ideal Conversation Flow. In the case where the user deviates from this ideal path of the conversation flow, the chatbot can fall back to several predefined responses to prompt the user to answer a question again (e.g. "Sorry, I didn't understand you... Please answer the question above by using the buttons"). To ensure that participants answer the open-ended probe questions comprehensively, a minimum answer length of ten characters was stipulated, with the chatbot prompting the user to elaborate on their answer in case this minimum was not met.

3.2 Personality Manipulation

In line with previous research on behaviour cues associated with *human* personality, we systematically manipulated the chatbots' language to imbue them with different human-like personalities. To this end, we first wrote the conversation flow for the average chatbot as a baseline and then adapted the responses for the introverted and extraverted chatbots as described in the following. Figure 1 illustrates extracts from the conversation as expressed in the three chatbots, whilst the complete conversation flows for all three chatbots can be found in the Supplementary Material.

To allow participants to better distinguish and recall the three chatbots for the final evaluation, all chatbots introduce themselves by name. We chose female names for the chatbots (Emily, Diane, Isabel) because users are most familiar with female characters for conversational agents. As personality markers are partly dependent on gender [55], we only incorporate behaviour cues associated with high or low Extraversion for females or across genders.

3.2.1 Average Chatbot. The average chatbot is designed to exhibit a medium level of Extraversion, which corresponds to the average Extraversion level in human populations [33, 113]. Little is known about the language use associated with an average level of Extraversion because previous research usually demonstrates positive or negative correlations between verbal cues and Extraversion [54, 94]. Hence, this chatbot does not implement any cues associated with higher or lower levels of Extraversion and uses the original, unmodified DISE questions [4]. Furthermore, the average chatbot is friendly and polite but does not disclose any personal information about itself or react in any emotional way to the user. For presenting the user with their stress feedback, this chatbot uses neutral, grey charts and line charts with points as people who are average in terms of Extraversion were found to prefer line charts with points (vs line charts without points) [5].

3.2.2 Extraverted Chatbot. To imitate an extraverted personality, the extraverted chatbot is enthusiastic and expresses its delight to communicate with the user (e.g. "Perfecttt 💪 thanks so much for your answers, *username*! That's all for today, I'll talk to you tomorrow. Already looking forward to it! 😍"). Due to an extravert's sociable nature, this chatbot mentions social interests such as meeting other chatbots [54, 131] and tends to jump from topic to topic [44]. To build a personal relationship with the user, this chatbot makes frequent use of the user's name. On the other hand, the extraverted chatbot is also assertive and commanding as characteristic for extraverted people [1] (e.g. "It will be best for you to look at it and think about your last few days.").

Due to extraverted people's predisposition of being talkative and, thus, having a high total verbal output [37, 44, 55, 80], the extraverted chatbot writes long text messages with a bigger number of words than its average and introverted counterparts. Furthermore, the extraverted chatbot's language is informal and loose, for example, by using phrases such as "let's catch up", "take care", or "Yeay" along with fillers such as "so" or "anyways" [37, 44]. As characteristic for extraverted people, this chatbot draws on a narrow vocabulary with frequent words [37, 44] (e.g. "Now, think of the most stressful happening of this sort. Who was the person you didn't wanna argue with?").

The extraverted chatbot uses many positively valenced words, such as "love", "happy", or "perfect" [44, 94, 131], reflecting extraverted people's inclination to be optimistic. Furthermore, this chatbot repeatedly incorporates the phrase "looking forward to" [44], personal pronouns, in particular the first person singular pronoun [44, 55] (e.g. "I see that your answer was very short... Could you tell me a little bit more, so I can better understand how you feel?"), and swear words such as "damn" [80] as informed by prior research. To enrich the chatbot's communication with nonverbal cues, we included emojis associated with high Extraversion, such as 😋, 😎, 👋, ❤️, 😍 [73, 123]. Moreover, we manipulated the chatbot's writing by integrating multiple exclamation marks (e.g. "Hi *username*!") [44] and word expansions (e.g. "perfecttt") [55].

Although not common practice in Psychology, we slightly modified single words and added subordinate clauses to the DISE questions to account for the different chatbot personalities. These modifications were because we were interested in participants' perception of the chatbot personalities rather than stressor elicitation and because the DISE questionnaire was originally proposed as an interview guideline. In particular, the extraverted chatbot adds subordinate clauses, which increases the text length and adjusts the language to seem more engaged (e.g. "Since this time yesterday, did anything happen at work or school that would be stressful for

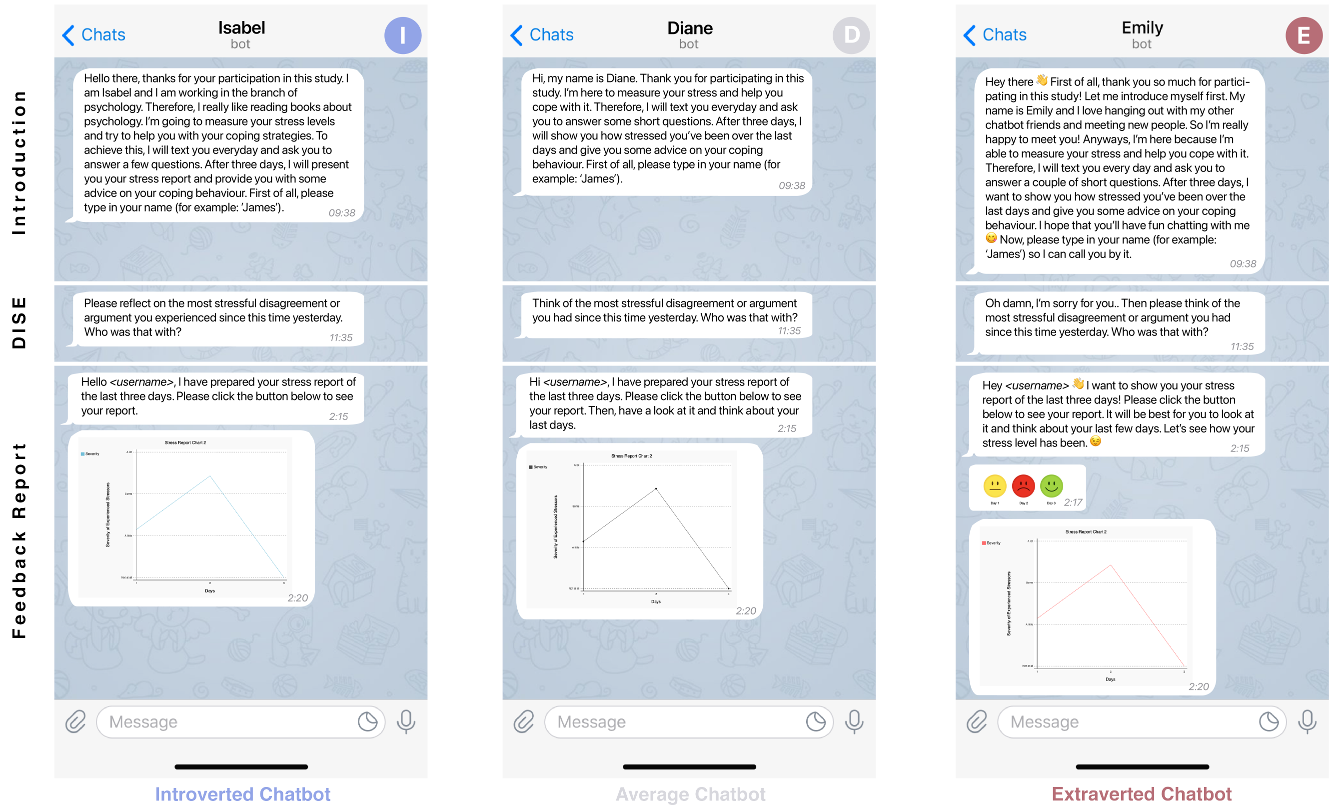

Figure 1: Excerpts from the conversation with the three chatbots for the introduction part (top), the DISE questionnaire (middle), and the feedback report (bottom) as expressed in the three chatbots (introverted left, average middle, extraverted right) to systematically imbue the chatbots with personality. Please note that for most messages, the introverted chatbot wrote fewer words than the average one but that sometimes the introverted chatbot's formal language caused slightly longer messages. The complete conversation flow can be found in the Supplementary Material.

most people? Of course, I mean other than what you've already mentioned. You don't have to tell me anything twice 😉").

For sending the user feedback about their stress level, the extraverted chatbot uses red-coloured charts in line with extraverts' predilection for exciting colours [63]. Because it was found that extraverted people prefer line charts without points [5], a simple version was used for this chatbot. In contrast to the other two chatbots, this chatbot also showcases a visual metaphor in the form of coloured emojis to illustrate a summary of the user's daily stress level, which was found to be enjoyed by extraverted people [109].

3.2.3 Introverted Chatbot. Conversely, the introverted chatbot is more reserved by disclosing only limited personal information and showing less emotional responses than the extraverted chatbot [44, 94]. Despite this, we let the introverted chatbot briefly introduce itself in the first message so as to convey interests associated with Introversion such as reading books. In line with introverted personalities [75], this chatbot introduces itself as being interested in psychology and reading books.

Regarding the speech style, the introverted chatbot writes more formally (e.g. "hello") and, thus, without word contractions (e.g. "I am" instead of "I'm") [43, 44]. Moreover, it uses a broader vocabulary with more infrequent words (e.g. "transpire", "reflect", "dispute") [37, 38, 44]. This chatbot's answers are shorter [24, 25, 37] and show a higher noun-to-verb ratio [37, 43] (e.g. "thanks for your participation in this study").

Due to the link between human personality and choice of words, the introverted chatbot is equipped with words related to achievement and work (e.g. "To achieve this, I will text you every day.") [131] and quantifiers such as "a few" [44]. In contrast to the extraverted version, this chatbot does not adorn its messages with any emojis as we did not find a negative correlation between Extraversion and emojis that would have been meaningful in this context.

We also slightly adjusted the DISE questions for the introverted chatbot. In particular, this chatbot employs a more infrequent, formal vocabulary, such as "transpire" or "occur" instead of "happen", "encounter" instead of "experience", or "reflect" instead of "think" (e.g. "What transpired and why did you decide not to get into an argument about it?") [37, 44].

For providing the user with their feedback report, the introverted chatbot showcases blue-coloured feedback charts as introverts tend to prefer calm colours that reduce excitement [63]. Similar to the extraverted chatbot, this chatbot also sends line charts without points [5]. Due to introverted people's inclination to use goal-oriented and efficient language [53], this chatbot asks the user to set goals after reading their feedback report.

3.3 Implementation

We developed the chatbots for the instant messaging application Telegram[2] to support a platform-independent solution integrated in a messenger app familiar to the participants. In Telegram, bots appear just like users but are controlled by their developers via the secure HTTPS Bot API. The chatbots ran on university servers and answers were saved in a database.

Supporting a smooth conversation, the chatbots used a combination of closed single-choice questions, implemented as inline keyboard buttons, and open questions. Based on the current state of the conversation flow (cf. Section 3.1), the chatbot selected the next text block or question from a dictionary that included all predefined text elements for the respective chatbot version. For the feedback report, we automatically generated bar and line charts to illustrate participants' stressors. In addition, we conducted a sentiment analysis on users' free-text answers, using the Python Natural Language Toolkit VADER [16], which employs a simple lexicon-based sentiment model.

4 RESEARCH DESIGN

To examine (1) whether human personality markers in language can be used to imbue chatbots with different levels of Extraversion, (2) users' preference for these different chatbot versions, and (3) the influence of user personality on this preference, we conducted a within-groups field study for 12 days. In particular, we asked $N = 34$ participants to interact with the three chatbot versions once a day for four days each on their personal phones. We counterbalanced the order of the three chatbot versions using a Latin Square.

4.1 Pre-Screening and Participant Recruitment

We recruited participants using *Prolific*[3]. To exclude language proficiency and dialect as confounding factors, we only distributed our study among British English first language speakers.

Because one of the study's goals is to examine the relationship between a user's personality and their preference for a chatbot personality, we pre-screened participants to ensure variability in our sample's self-reported Extraversion scores, which we collected via the corresponding twelve items in the established Big Five Inventory-2 questionnaire (BFI-2) [113]. Based on the mean population score in Extraversion of 3.23 ($SD = 0.80$) [113] on a scale from 1 (minimum) to 5 (maximum), we divided participants into three groups: introverted ($M - 1SD$: < 2.44), average ($M \pm 1SD$: $2.44 - 4.03$), and extraverted ($M + 1SD$: > 4.03). We estimated the pre-screening survey to take three minutes and, thus, offered £ 0.38 as compensation. We randomly invited participants for each of the three groups to join our main study. Until all three groups consisted of at least ten participants, we recruited a total of $N = 181$ participants. Notably, only few participants classified as extraverted answered the pre-screening questionnaire.

Upon invitation after the pre-screening questionnaire, $N = 34$ participants (12 extraverted, 10 average, 12 introverted) joined the main study. Based on several pilot runs, we estimated participants to spend approximately 120 minutes on the main study, resulting in a compensation of £ 15.00.

4.2 Procedure

Participants started the study by completing an online questionnaire. Afterwards, they were directed to the Telegram chatbot. On the first day of interaction with each chatbot, the chatbot introduced itself and asked the participant to answer the DISE questions as outlined in Section 3.1. Participants were prompted with a push-notification to answer the DISE questionnaire again on the second and third interaction days. On the fourth day, the chatbot showed participants their stress feedback report and sent a link to another online questionnaire in which participants rated the chatbot's Extraversion, its usability, and their desire to interact with it again. The procedure was then repeated for the other two chatbot versions. On the final (12th) day, participants completed a post-interaction questionnaire. Figure 2 shows an overview of our study design.

4.3 Introduction Questionnaire

In the first questionnaire, participants were introduced to the study purpose and procedure as well as asked for their consent in line with our institution's regulations. Henceforth, we collected participants' self-reported personality traits via the 15 items of the Big Five Inventory-2 questionnaire (BFI-2-XS) [114] and demographic information. To merge their data, participants were instructed to generate a personal ID, which was then ascertained in their interaction with the Telegram chatbot and all subsequent questionnaires.

4.4 Chatbot Interaction and Measurements

Participants interacted with each of the three chatbot versions for four days as described in Section 3.1. Upon receiving their feedback report, the chatbot presented participants with a link to an online survey. In this online survey, participants were asked to describe their perception of the chatbot's personality and rate their interaction desire as well as the chatbot's usability. To ensure that participants completed the online evaluation survey, they received a code word, which was validated by the chatbot.

4.4.1 Perceived Personality. First, we instructed participants to describe their impression of the chatbot's personality in their own words in a free-text field. By collecting participants' open description before completing the questionnaire, we were interested in the salience of personality traits and differences in perceptions that cannot be captured by means of a questionnaire [67, 126]. Second, we asked participants to indicate the chatbot's level of Extraversion by filling out the twelve corresponding BFI-2 items [113].

4.4.2 Desire to Interact. To evaluate their interaction experience, participants were asked to indicate on a five-point Likert scale how much they would like to interact with the respective chatbot again in the future.

[2]https://telegram.org, last accessed April 10, 2022
[3]https://www.prolific.co/, last accessed April 10, 2022

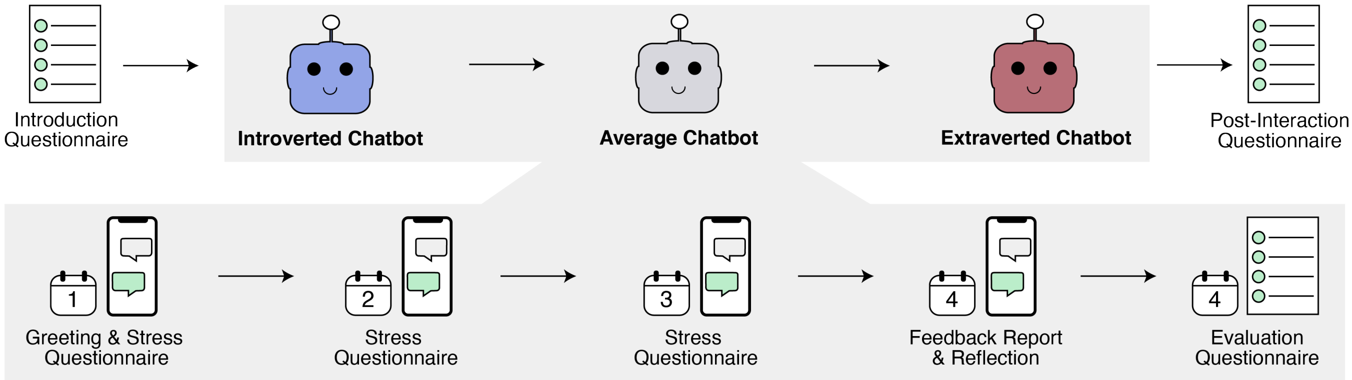

Figure 2: Research design: Participants interacted with the three personality-imbued chatbots for four days each (three days of tracking stress and the fourth day for feedback and reflection). On the fourth day, participants evaluated the respective chatbot in an online questionnaire. Upon completing the study with all three chatbots, participants compared the three versions in a final questionnaire.

4.4.3 Usability. To exclude differences in usability as a confounding factor in users' preference for the three chatbot versions, we assessed the usability of the chatbots via the System Usability Scale (SUS) questionnaire [21]. As of today, there is no standardised instrument for measuring chatbot usability and user experience [59, 111]. We decided on the SUS as it has been used previously to evaluate the usability of conversational agents [36, 56] and provides a fast and short assessment.

4.5 Final Post-Interaction Questionnaire

After participants had interacted with all three chatbot versions, we asked them to complete a final questionnaire. To refresh their memories about the chatbots they had interacted with earlier and allow comparison, we first showed them a screenshot of each of the three chatbots' welcome message. Afterwards, participants were instructed to rank the three chatbots according to which one they liked best. Finally, we asked participants to briefly describe the reasons for their decision in a free-text field.

4.6 Participants

Our sample comprised $N = 34$ participants (70.6% self-identified as female, 29.4% as male, none non-binary or self-described). Participants were between 19 and 59 years old ($M = 32.71$, $SD = 10.54$) and predominantly highly educated, with 67.7% having a university degree, 26.5% an A-level degree, and 5.9% a middle school degree. The majority of participants were employed full time (44.1%), part time (20.6%), or were students (23.5%), with the remaining participants being self-employed or retired (11.7%).

The majority of participants (70.6%) had interacted with a chatbot before, although this interaction was rather rare (35.3% interact with chatbots once a year, 23.5% once a month, and 11.8% once a week). On a scale from 1 = *very bad* to 5 = *very good*, participants' median previous experience with chatbots was neutral (20.6% of participants), with 23.5% of participants rating their experiences as good and 26.4% as (very) bad, respectively.

5 RESULTS

5.1 Usability of the Chatbots

On a scale from *0 = low usability* to *100 = high usability*, all three chatbots achieved similarly high usability ratings on the SUS scale (extraverted chatbot: $M = 80.81$, $SD = 12.79$; average chatbot: $M = 79.71$, $SD = 14.77$; introverted chatbot: $M = 78.82$, $SD = 16.51$). These scores represent "good" usability [10] and are comparable to results in similar health application chatbots [58]. Because the data were not normally distributed ($W = 0.93$, $p < .001$), we conducted a Friedman test, which did not yield any significant effect of the chatbot on the SUS usability score ($\chi^2(2) = 0.98$, $p = .611$). These results indicate that differences in users' preference for the chatbots are likely not caused by their usability.

5.2 Personality Questionnaire Assessment

In line with the BFI-2 personality questionnaire instructions [113], we calculated the mean Extraversion score for each of the three chatbots (scale from 1 = *introverted* to 5 = *extraverted*). As shown in Figure 3, participants perceived the extraverted chatbot on average as more extraverted ($M = 4.01$, $SD = 0.54$) than the average ($M = 3.41$, $SD = 0.72$) and introverted chatbots ($M = 3.32$, $SD = 0.62$). For comparison, the population mean score for human Extraversion is 3.23 ($SD = 0.80$) [113]. Whilst the manipulation can be considered successful for the extraverted and average chatbots, the introverted chatbot was not perceived as introverted as expected. A repeated-measures ANOVA supports these findings ($F(2, 66) = 16.23$, $p < .001$, $\eta_p^2 = 0.33$). Pairwise post-hoc tests using Bonferroni correction for p-value adjustment revealed significant differences between the extraverted and introverted chatbots ($p < .001$, $d = 1.18$) and the extraverted and average chatbots ($p < .001$, $d = 0.94$), whereas there was no significant difference between the introverted and average chatbots ($p = 1.00$, $d = -0.13$). Despite not being perceived as introverted as intended, we continue to refer to the chatbot with the lowest level of Extraversion as the *introverted* chatbot as it was designed as such and for clarity. We will address alternative personality perceptions in the Discussion.

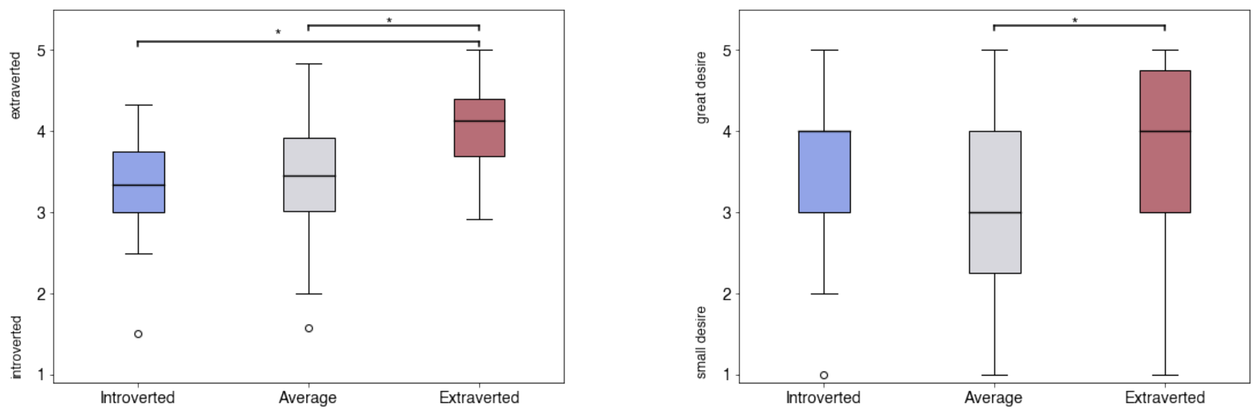

Figure 3: Participants' perception of the three chatbots' level of Extraversion using the BFI-2 personality questionnaire [113] (left) and desire to interact with the three chatbots again (right).

Figure 4 illustrates that the items in the personality questionnaire differ with regard to the success of the personality manipulation. Whilst evaluation of the chatbots' Extraversion was ordered as intended for the majority of the items, participants did not perceive a meaningful difference among the three chatbots for *"has an assertive personality"*, *"is less active than other people"*, *"finds it hard to influence people"*, and *"prefers to have others take charge"*.

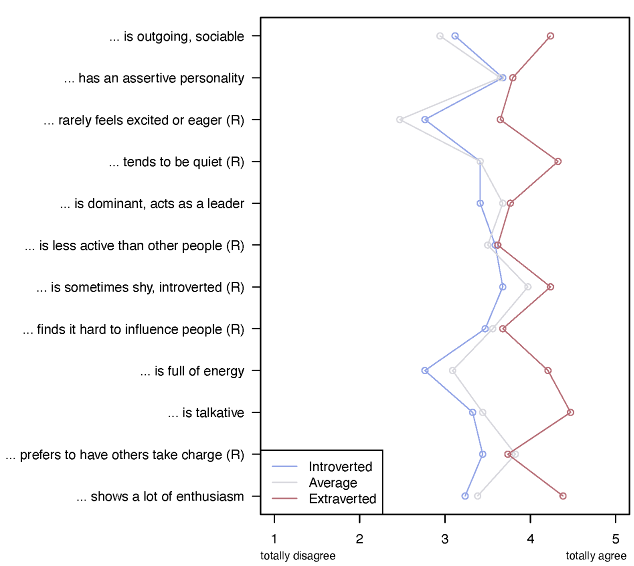

Figure 4: Participants' perception of the three chatbots' level of Extraversion on the single items in the BFI-2 personality questionnaire. Items denoted by an "(R)" are reverse-keyed; that is, the score displayed for these items is the opposite of the original score to allow for easier comparison among the three chatbots.

5.3 Open Personality Descriptors

We collected participants' open descriptions of the chatbots' personalities in a free-text field before they rated the chatbot on the personality questionnaire items. From these descriptions, we extracted mainly adjectives as commonly used in personality psychology [34, 47, 87] and individual phrases characterising the chatbots' personalities, which we henceforth refer to as *descriptors*. We excluded descriptors from the analysis that are over-evaluative (e.g. "great"), refer to demographics (e.g. "young"), or outline the lack of a characteristic (e.g. "less inquisitive").

Afterwards, we assigned each of the resulting 160 unique descriptors to one of the Big Five dimensions and corresponding sub-clusters (cf. Table 1). For this mapping, we referred to Goldberg's list of 339 personality adjectives [48], which constructed the basis of the Big Five model and was used in related work on conversational and embodied virtual agent personality [67, 126]. For example, participants described the extraverted chatbot as "cheerful", which belongs to the Big Five dimension Extraversion in the cluster *Optimism* [48].

For all descriptors not included in Goldberg's list (62.5%), we systematically searched for synonyms as follows. We first looked up the descriptor in the *Oxford English Dictionary*[4] (OED), collecting synonyms from the word's definition or thesaurus (33.1% of all 160 unique descriptors). We then matched these synonyms with Goldberg's list of personality adjectives [48]. For example, for participant's term "bubbly", the OED provided the definition *"Of a person, or his or her personality, nature, etc.: vivacious, full of high spirits"*.[5] As "vivacious" is included in the cluster *Spirit* in high Extraversion, we also assigned "bubbly" to this cluster. In the case of multiple definitions for a descriptor or synonyms that belong to different clusters, two authors extensively discussed the options and decided on the most likely interpretation (e.g. "upbeat" has

[4]https://www.oed.com, last accessed April 10, 2022
[5]https://www.oed.com/view/Entry/24083?redirectedFrom=bubbly, last accessed April 10, 2022

the synonyms "cheerful" and "optimistic", which map the Extraversion cluster *Optimism*, whereas another synonym, "vigorous", maps the Extraversion cluster *Energy level*[6]). For 6.3%, we did not find a synonym in the OED and instead turned to the synonym website *Thesaurus*[7] following the same procedure as for the OED.

Table 1 in the Appendix summarises participants' descriptors, which were assigned to the Big Five dimensions. Please note that the counts refer to how often descriptors in one cluster were mentioned across all participants. That is, if one participant characterised the chatbot as extraverted with several descriptors, then each descriptor will add up individually as we found the number and variety of descriptors used for one trait more meaningful in the context of personality perceptions than how many participants included at least one descriptor for each dimension. As illustrated in Table 1, participants mentioned descriptors for high Extraversion 20 times for the extraverted chatbot, in particular highlighting its *Optimism*. In contrast, descriptors for high Extraversion were mentioned only four times for the introverted and ten times for the average chatbot. Notably, the majority of these descriptors refer to the chatbots' *Candour*, which may have resulted from the average chatbot using less but more concrete and explicit words than the more verbose extraverted version. On the other hand, five descriptors relating to low Extraversion were used for the introverted chatbot, primarily pointing out its *Reserve*, and four for the average one.

Based on the categorisation of participants' open descriptors into the Big Five personality traits, we calculated a personality dimension score for each chatbot by subtracting the number of low level descriptors from the number of high level descriptors (e.g. Extraversion score for the introverted chatbot: 4 high level - 5 low level = -1). These scores are visualised in Figure 5. We used the biggest count (that is, 42 in Agreeableness for the extraverted chatbot) as maximum and minimum, respectively, for all dimensions. Prominently, the extraverted chatbot was perceived as more extraverted than the other chatbots, which were perceived as less different from one another, underpinning the personality questionnaire results. However, the chatbots' Agreeableness was most salient to participants, who particularly highlighted their *Amiability*. The reasons as to why the three chatbots were all perceived as rather agreeable differed. The extraverted chatbot was primarily perceived as natural (e.g. "informal", "casual", "relaxed") and cooperative ("helpful", "personable), whilst the *Empathy* cluster was emphasised for the average ("nice", "kind", "thoughtful") and introverted ("understanding", "thoughtful", "sympathetic", "kind") chatbots.

Conversely, attributes associated with high Conscientiousness were most noteworthy to participants for the introverted and average chatbots. Both chatbots were perceived as high in Efficiency (e.g. "professional", "straight to the point") and the introverted chatbot's "formal" demeanour was stressed (see Table 1). Whilst the extraverted and introverted chatbots were perceived as higher in Openness than the average chatbot due to their observed inquisitiveness and interest, descriptors falling into the Neuroticism dimension were barely mentioned.

Some descriptors (14.4%) did not match any of the Big Five dimensions. Instead, we inductively grouped these descriptors into two clusters. The first cluster-pair *Usable - Unusable* comprises descriptors such as "informative" or "dysfunctional", referring to the chatbots' usability and task compliance. Positive evaluations of the chatbots' usability were mentioned slightly more often for the introverted and average chatbots. The second cluster-pair compares the chatbots' perceived *Humanness* and *Thingness*, which we subsumed under *Artificiality*. Notably, *Thingness* emerged more frequently for the average chatbot, as illustrated by descriptors such as "robotic", "unnatural", "clinical", and "no personality". Finally, 8.8% of the descriptors could not be assigned to any of the aforementioned dimensions and were included in an *Others* category (cf. Table 1).

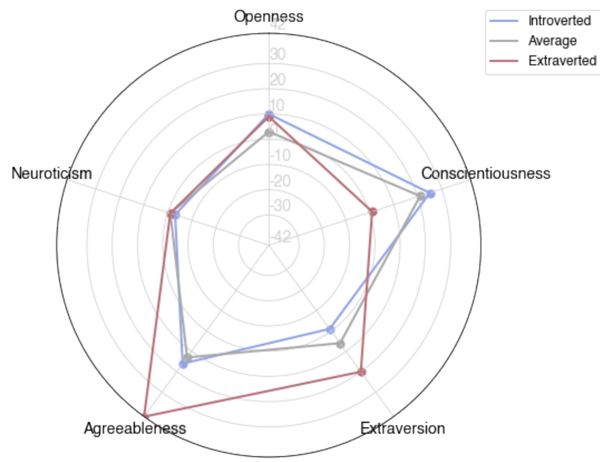

Figure 5: Based on participants' open personality descriptors and their assignment to the Big Five, we calculated a personality dimension score for each chatbot by subtracting the number of low pole descriptors from the number of high pole descriptors.

5.4 Preference to Interact with the Chatbots

5.4.1 Desire to Interact with the Chatbots. After participants had used a chatbot for four days, we asked them in an online survey whether they would like to interact with this chatbot again in the future. Figure 3 shows participants' evaluation of the three chatbots. On a scale from 1 = *strongly disagree* to 5 = *strongly agree*, participants expressed on average a greater desire to interact with the extraverted chatbot (Mdn = 4, M = 3.71, SD = 1.12) than with the introverted (Mdn = 4, M = 3.47, SD = 0.96) and average chatbots (Mdn = 3, M = 3.15, SD = 1.13). Because the data were not normally distributed (W = .900, p < .001), we conducted a Friedman test, which determined a significant effect of the chatbot on participants' interaction desire ($\chi^2(2) = 6.64$, p = .0361). Pairwise Wilcoxon post-hoc tests using Bonferroni correction yielded a significant difference between the desire to interact with the extraverted and the average chatbots (p = .046, r = 0.42) but not between the other two pairs (extraverted vs introverted: p = .813, r = 0.25; introverted vs average: p = .202, r = 0.25).

[6] https://www.oed.com/view/Entry/219826?rskey=AVnMMs&result=2, last accessed April 10, 2022
[7] https://www.thesaurus.com, last accessed April 10, 2022

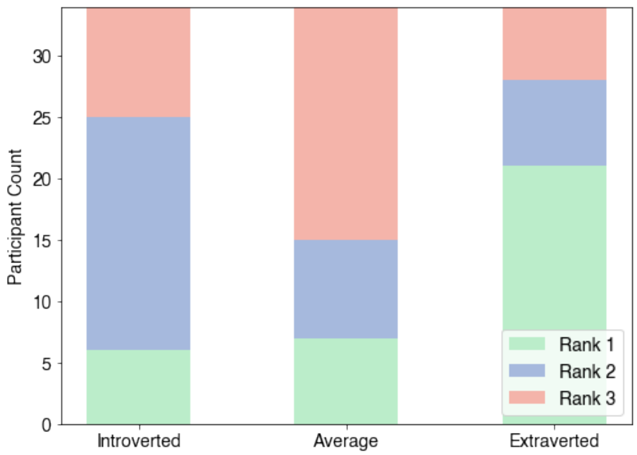

Figure 6: Participants' final ranking of the three chatbots according to which one they liked best after interacting with all three chatbots for four days each.

5.4.2 Ranking of the Chatbots. Participants' preference for the three chatbot versions is also reflected in their final ranking after they had interacted with all three chatbots. As shown in Figure 6, the extraverted chatbot was overall ranked as first choice by 61.8% of participants, whereas the introverted chatbot was ranked second place by the majority of participants (55.9%) and the average chatbot third place (55.9%). We performed a Friedman test on the rankings and found a significant difference between the rankings ($\chi^2(2) = 11.12, p = .004$). Conducting Wilcoxon signed-rank tests as post-hoc tests with Holm correction applied, we found that the extraverted chatbot was significantly more preferred over both the introverted ($p = .048, r = 0.39$) and the average chatbot ($p = .010, r = 0.50$). However, there was no difference between the introverted and average chatbots ($p = .226, r = 0.19$).

5.5 Reasons for Ranking the Chatbots

After ranking the chatbots, we asked participants to briefly describe the reasons for their rankings. We then inductively coded and clustered these reasons.

5.5.1 Reasons for Ranking the Extraverted Chatbot. Of the 21 participants who ranked the extraverted chatbot first, 14 gave as reason for their ranking that this chatbot was "the friendliest" (e.g. P4). Participants appreciated the use of emojis (P3, P11, P18, P31), its humanness (P3, P4, P6, P12), which felt "closer to speaking to a real person" (P4), and that the chatbot disclosed more information about itself (P22). Moreover, participants highlighted that the extraverted chatbot was "enthuasiastic" (P7, P15), "informal" (P12, P30), "happy" (P1), "chatty" (P3), "warm" (P7), "fun" (P15), "interesting" (P23), "pleasant to talk to" (P33), "more on the level" (P34), "supportive", and "seemed like she genuinely cared more" (P24), which "feels more like a friendship with someone who checks in on you" (P1).

On the other hand, participants who did not rank the extraverted chatbot first stressed that they did not like the chatbot's emoji use (P8, P19, P26, P27), with participants finding the chatbot "too informal" or "casual" (P26, P27, P32), "overfamiliar" (P19, P32), or "unprofessional" (P32). Moreover, participants criticised the chatbot for being "annoying" (P5, P8, P17, P26), "over-friendly" (P10), "too young" (P8), "trying to be cool" (P9), or being "excessively enthusiastic" (P26).

5.5.2 Reasons for Ranking the Average Chatbot. Of the seven participants who found the average chatbot to be the best, three cited the friendliness of this bot as the reason. Two participants emphasised that the chatbot was "straightforward" (P17, P32). Furthermore, participants enjoyed the chatbot being "supportive" (P8), "engaging" (P17), "professional" (P26), "inquisitive" (P28), "to the point" (P32), and highlighted its "plain" language (P9).

Conversely, participants who did not rank this chatbot first particularly pointed to the chatbot's artificiality, describing it as "robotic" (P12, P20, P34), "mechanic" (P6, P7), "cold" (P7, P20), "clinical" (P34), "impersonal" (P7), "black and white" (P15), "generic" (P6), "neutral" (P15), and without "much personality" (P33). Furthermore, participants criticised that the chatbot was "not fun" (P3), "not likable" (P3), "distant" (P10), "focused" (P12, P14), "unfriendly" (P10, P23), "uninterested" (P23), and "unprofessional" (P23), and did not "give any information about herself" (P7, P13).

5.5.3 Reasons for Ranking the Introverted Chatbot. Six participants placed the introverted chatbot first. Two participants found the chatbot "friendly" (P10, P19) whilst also being "professional" (P10, P29) and "informative" (P19, P27). Moreover, participants highlighted that the chatbot was "studious" and "knowledgeable" (P25), used a "clear language" (P27), and provided appropriate "information about herself" (P13). P29 liked "the personality coming through whilst remaining professional".

However, some participants did not like these characteristics but found the introverted chatbot too "formal" (P26), "serious" (P12), "annoying" (P5, P9), "overbearing" (P21), "boring" (P26), "not easy to react to" (P8), and "complicated" (P17, P26), describing it as neither "fun" (P3), "likable" (P3), "friendly" (P26) nor "approachable" (P21). P34 compared the chatbot to *Microsoft Clippy* and P22 would have liked to read more information about the chatbot.

5.6 Engagement with the Chatbots

Apart from participants' subjective rating of the three chatbots, we also analysed the total number of words in participants' free-text responses in order to gauge user engagement. We chose this metric as it has been used in comparable previous work on user-chatbot interaction [51, 66, 130], indicating that engaged users are more willing to write longer responses [130].

We first inspected how many stressors participants experienced during the use of each chatbot (i.e. number of positively answered stressor questions). From a total of 21 possible stressors per chatbot per participant (3 days × 7 stressor questions), participants affirmed on average $M = 1.61$ questions ($SD = 1.12$) to the extraverted chatbot, $M = 1.64$ questions ($SD = 1.14$) to the average chatbot, and $M = 1.61$ questions ($SD = 1.12$) to the introverted chatbot. Thus, participants responded equally frequent to the stressor questions regardless the chatbot version ($F(2, 64) = 0.01, p = .990, \eta_p^2 < 0.001$). Thus, with this analysis, we can rule out the possibility that any effects reported below would be due to participants systematically answering more questions for certain chatbot versions.

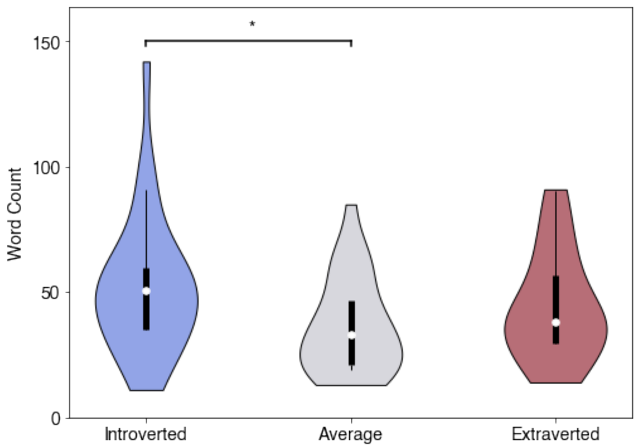

Figure 7: Participants' average word count to describe a single stressor in the interaction with the three chatbots.

Based on this finding, we explored whether the amount of text that participants wrote in response to the chatbots' open-ended questions (i.e. the number of words to describe the stressor) varied depending on the Extraversion level of the respective chatbots. To do so, we concatenated participants' answers for all probe questions belonging to a stressor question. As a reminder, for a positively answered stressor question (a stressor), the participants were asked to describe this stressor by responding to several probe questions. Because the total number of words a participant writes to a chatbot is dependent on the number of stressor questions answered, we first computed the mean number of words a participant has used over all positively answered stressor questions per day. Afterwards, we calculated how many words participants wrote on average to each chatbot to describe a stressor over the duration of the three stress tracking days. Please note that we did not exclude stop words from our word count analysis as we were interested in the total number of words independent of their meaningfulness. In particular, we expected that participants use more words, including meaningless fillers, to converse with the verbose extraverted chatbot as people are known to match their interlocutor's language style [86].

To account for missing values because not all participants reported stressors every day, we ran a linear mixed model (random-intercept-fixed-slope) to investigate whether the word count depended on the respective chatbot version (n = 83 observations nested in 33 participants; that is, one participant did not experience any stressors over the 12 interaction days). The chatbot version was dummy-coded and the average version served as the reference category. The estimated word count for answering the average chatbot's probe questions was $M = 36.33$ per stressor (cf. Figure 7). Comparing the extraverted to the average chatbot, participants, on average, did not significantly use more words ($b = 8.53$, $SE_b = 4.98$, $t(55.46) = 1.71$, $p = .093$). However, when contrasting the introverted to the average chatbot, participants' word count increase was significant ($b = 16.11$, $SE_b = 4.96$, $t(54.27) = 3.25$, $p = .002$). A comprehensive results table with all model estimators can be found in the Supplementary Material.

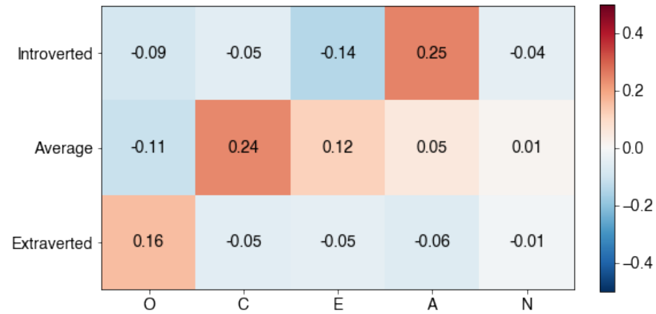

Figure 8: Spearman correlations between participants' own personality traits (OCEAN) and their desire to interact with the different chatbot versions. The colour of the squares indicates the height and direction of the correlations.

5.7 Relationship between User Personality and Desire to Interact with the Chatbots

In a next step, we explored if the participants' own personality was related to their preference for different chatbot versions. We descriptively calculated Spearman correlations between participants' Big Five traits (for each dimension) and their desire to interact with the three chatbots (see Figure 8). In the majority of cases, we found no or only very small correlations, with the exception of two small to medium correlations: more conscientious participants had on average a greater desire to interact with the average chatbot ($r = 0.24$) and more agreeable participants had on average a greater desire to interact with the introverted chatbot ($r = 0.25$). To account for the repeated measures structure in our data, we ran a linear mixed model (random-intercept-fixed-slope) to explore whether participants' desire to interact with the chatbots was related to their own Big Five personality traits. However, we could not find any significant effect of participants' traits on their desire to interact (all $p > .05$). A comprehensive results table with all model estimators can be found in the Supplementary Material.

6 LIMITATIONS

Our data, method, and findings are limited in several ways and should be understood with these limitations in mind. Although our within-subjects design had the advantage to parallelise person-specific confounding variables, it might have fostered contrast effects in participants' Extraversion perception between the chatbot versions. If users are only presented with a single chatbot version (e.g. corresponding to their own personality), then their perception of the chatbot's personality may differ [67]. In addition, the fact that participants were repeatedly presented with the same questions may have caused response fatigue and biases over time (e.g. personality inventory items, DISE questions).

Other limitations of our study relate to the sample. First, due to our strict recruiting criteria, we were only able to capture a relatively small sample of 34 participants. A sensitivity power analysis revealed a minimum detectable effect size of $\eta_p^2 = 0.05$ (with $1 - \beta = .85$, $\alpha = .05$). In other words, with our sample we could find at most medium effects, but no small effects, which are quite common in psychological research contexts [101]. Second, our sample was highly educated and restricted to British English first language speakers. Thus, socio-demographic or cultural effects could pose a problem for generalising results to other populations.

Finally, due to the innovative nature of our project, we considered the formulation of guiding research questions more adequate than following concrete hypotheses. We would therefore like to point out to the reader that future work is needed to replicate our results and take an in-depth look into the discovered effects.

7 DISCUSSION

7.1 RQ1: Manipulating Chatbot Personality

Paying tribute to the continuous rather than dichotomous nature of personality traits, we intended to manipulate three levels of Extraversion and evaluated the full range of participants' perceptions using a standardised personality inventory and open descriptions.

7.1.1 Reflections on Synthesising Extraversion. Whilst the personality manipulation can be considered successful for the extraverted chatbot, transferring verbal cues from human behaviour was not sufficient to create a truly introverted chatbot. Our findings demonstrate that it is difficult to differentiate between different levels on the lower end of the Extraversion spectrum by means of verbal cues alone. This result is evident by the perceived similarity of the introverted and average profiles and the difference to the extraverted one, both in the inventory results and the open personality descriptions.

Although participants identified the behaviour cues as intended for the introverted chatbot (e.g. formal language, conciseness), these characteristics were attributed to the chatbot's Conscientiousness instead of Introversion, which was also observed in the open personality descriptions by Ruane et al. [103]. A reason why lower levels of Extraversion were difficult to synthesise could be that we kept several interaction variables constant across the three chatbots, such as the number of interactions and notifications. In particular, a "truly introverted chatbot" might not prompt the user to fill out the daily questionnaire but only reacts if the user initiates the conversation. Future work should evaluate the effect of these interaction cues on personality perception; however, conversation designers then have to make a trade-off between personality perception and the application's task description; that is, for example, completing a daily questionnaire.

Albeit the problem of manipulating extreme expressions of personality in CAs has also been acknowledged for voice interfaces [82, 125] and embodied virtual agents [67], replenishing agent behaviour with non-verbal cues is likely to mitigate the effects. Because non-verbal behaviour tends to eclipse verbal cues [9], personality markers, such as avoiding gaze or talking quietly, are probably more powerful to convey Introversion. As personality traits were found to be associated with the use of emojis [73, 123], which are commonly used as a surrogate for non-verbal cues in text messaging [127], future work could also examine the relationship between the perception of low level traits in chatbots and their emoji usage.

7.1.2 Reflections on Perceived Personalities. Taking together the results from the Extraversion inventory and participants' open descriptions, we reflect on the three chatbots' personalities as perceived by our participants using the Big Five.

Participants perceived the *extraverted chatbot* as such in the inventory and highlighted its cheerful, upbeat nature and its enthusiastic spirits. The most salient characteristics for participants were on the Agreeableness dimension, stressing the friendliness, cooperation, and casual demeanour. Furthermore, participants noted its interest (high Openness) whilst the chatbot's levels of Conscientiousness and Neuroticism were impertinent. Summing up, this chatbot was perceived as agreeable, extraverted, and open.

The *average chatbot* was rated as having average Extraversion levels, primarily manifested through its assertive and direct attitude. Most salient to participants was the chatbot's high Conscientiousness epitomised through being very straight to the point, focused, and formal. Whilst some participants highlighted its friendly, kind, and helpful demeanour (high Agreeableness), others perceived it as cold and impersonal. In contrast, its Openness and Neuroticism were rather inconspicuous to participants. Moreover, this chatbot was often discerned as unnatural and robotic although a human with an average level of Extraversion is less likely to be perceived as such. An average level in a personality dimension is most likely achieved if no characteristics on this dimension are salient to the rater (e.g. in the Big Five Inventory, a rater receives an average rating by always choosing the middle answer option "neutral; no opinion"). Therefore, it is more likely that the chatbot's efficiency and lack of disclosing personal information caused the perception of being artificial. Summing up, this chatbot was perceived as conscientious, agreeable, artificial, and averagely extraverted.

Participants also ascribed average Extraversion levels to the *introverted chatbot*, with its reverse and detached attitude being mentioned in the open personality descriptions. The most salient characteristic was again Conscientiousness due to its formal, professional, straight to the point, focused, and persistent behaviour. Apart from its friendliness and helpfulness, participants highlighted this chatbot's empathic, understanding, and helpful demeanour on the Agreeableness dimension. Furthermore, this chatbot is also perceived as interested (high Openness), whilst few participants noticed a matter-of-fact demeanour (low Neuroticism). Summing up, this chatbot was perceived as conscientious, agreeable, open, and averagely extraverted.

7.2 RQ2: The Effect of Chatbot Personality on Preference and Behaviour

Our results signify that the perceived personality of a chatbot has an effect on user preferences and interaction behaviour, which we describe in this subsection.

7.2.1 Users Enjoy Interacting with Human-like Personalities. Although consumer reports and previous research have indicated that users enjoy interacting with CAs with human-like personalities [32, 121], there has been scepticism around the benefits this type of naturalness may produce [40] given that today's CAs often fall short of users' expectations [32, 70, 97, 100]. Our findings indicate that the majority of participants seems to enjoy interacting with a chatbot that exhibits a human-like personality in the context of a daily stress tracker, thereby supporting assumptions in previous work [39]. Several participants explicitly highlighted in the reasons for their ranking that with the extraverted chatbot it felt more like talking to an actual human, whereas the robotic and mechanic nature of the average chatbot was criticised.

In contrast to their ranking, our word count analysis showed that participants wrote significantly more text when talking to the introverted chatbot than when chatting with the average chatbot. This difference is particularly surprising, given that the chatbots' personality profiles were quite similar, both in the BFI-2 questionnaire and participants' open descriptions. The latter points to a difference in participants' perception of the chatbots' humanness, with the average chatbot being assigned several characteristics highlighting its artificiality. Hence, the chatbots' perceived humanness may determine not only subjective preferences for chatbot personality but also lexical choices, thereby echoing recent work by Doyle et al. [39] on dimensions of partner models for speech user interfaces.

7.2.2 The Role of Agreeableness in Personality Preference. Openly describing the chatbots' personality before completing the personality inventory allowed participants to mention those traits that they perceive as most salient and noteworthy [67]. For our chatbots, the predominant traits were clearly Agreeableness for the extraverted chatbot and Agreeableness and Conscientiousness for the average and introverted ones, confirming previous findings on open personality descriptions for chatbots [133], voice assistants [126], and virtual embodied agents [67]. As Extraversion and Agreeableness are also inter-correlated in human personality [49] due to their predisposition for social relationships and for sharing similar personality markers [80, 94], participants' perception of Agreeableness in particularly for the extraverted chatbot was to be expected. Hence, in line with prior work [67, 79], manipulating a single personality dimension might not be possible but, similar to human personality, personality manifestations have to be considered as multi-dimensional construct. CA developers should keep this in mind when designing and evaluating their CAs to ensure that the manipulation of one dimension does not have unintended consequences on another one.

Participants' reasons for ranking the chatbots corroborate the importance of Agreeableness with the extraverted chatbot's friendliness being repeatedly singled out by participants. Users' preference for Agreeableness in voice assistants [19] and chatbots [124] has been highlighted before. A reason for this preference could be that Agreeableness and Extraversion are particularly meaningful in interpersonal interaction [76, 83], which is relevant as users perceive CAs as helpful assistants [122]. Future work should examine the influence of these two personality traits more closely by untangling whether users' preference for this chatbot is caused by the perceived Extraversion, Agreeableness, or their interplay.

7.2.3 The Role of Introversion in User Interaction. Whilst our results show a subjective preference for the extraverted chatbot, participants disclosed more words to the introverted chatbot, thereby shedding further light on the impact of chatbot personality on user behaviour. Notably, Ruane et al. [103] found a similar effect, as participants subjectively preferred an extraverted and agreeable chatbot but had longer interaction times with the introverted, less agreeable one. This finding is surprising since previous work has shown that users tend to align their language to their artificial interlocutor [18, 31]. A reason for this reversed effect could be that participants took the introverted chatbot more seriously due to its professional and formal behaviour, thus putting more effort into the answer. Another explanation could be that the interaction with the extraverted chatbot was overall longer due to its verbosity so that participants felt less motivated to write long answers. Future work should examine if the chatbot's personality does not only influence the number of words written but also the response quality.

In summary, our results underline the significance of humanness, Agreeableness, and Extraversion in the context of designing chatbot personalities. In contrast to human personality, for which Extraversion stands out as having the closest link to observable behaviour [94], our findings emphasise the salience of Agreeableness, supporting prior work on identifying speech-based CA personality dimensions [126]. This adds much-needed information as to which personality traits CA designers should consider based on the application's goal when developing CA personalities.

7.3 RQ3: Individual Differences in Personality Preferences

Participants' rankings of the chatbots showed that the majority of participants preferred the extraverted chatbot. Participants' reasons, however, revealed that those characteristics of the extraverted chatbot that the participants who ranked this chatbot as their favourite particularly emphasise are precisely the reasons why the other participants rejected it. For example, the chatbot's "informal" and "friendly" nature were praised by some but described as "too informal" and "over-friendly" by others.

Despite these clear individual preferences, our results do not signify that this difference is caused by participants' own personality in contrast to previous work which found a similarity attraction effect between user Extraversion and preference for personalities of CAs [19, 82] or robots [8]. One possible reason for this lack of an effect could be our small sample. However, previous work also suggests that the similarity attraction relationship in human-agent interaction is a complex phenomenon. For example, Andrist et al. [8] only found a similarity attraction effect for introverted participants and their subjective preference for an introverted robot, none for extraverted participants. Potentially, the lack of an effect could also stem from the introverted chatbot not being perceived as introverted. Nass and Lee [82] showed a similarity attraction effect between user Extraversion and their preference for an extraverted voice but no similar effect for extraverted text.

Hence, our work has highlighted the advantage of including qualitative reasons for users' preferences to better understand individual needs. Although we could not find a relationship between participants' Extraversion and their preference for the average chatbot as expected, it was the favourite chatbot for 20.6% of participants. These results underline that users prefer different levels of Extraversion in chatbots. Future work should investigate more closely which user characteristics beyond personality (e.g. demographics or affinity for technology) are of prime importance for their preferences. In the case where only one version can be built, our findings suggest that the majority of participants find the extraverted chatbot most appealing for a stress tracking scenario. However, future work should examine if participants prefer other chatbot personalities in different contexts.

7.4 Reflections on CA Personality Assessment

As discussed before, the perception of a chatbot's artificiality seems to be detrimental to participants' subjective preference and engagement. However, these differences in humanness would not have been possible to decipher by means of a human personality questionnaire alone. The use of terms referring to the artificiality of a CA when openly asked to describe their personality has been observed in several studies [67, 95, 126, 133]. The importance of artificiality to characterise a CA personality has also been highlighted in our prior work on personality dimensions for voice assistants, which brought to light a personality dimension of *Artificiality* [126]. Similarly, Doyle et al. [39] identified human-likeness as a key dimension for users' partner models for speech agents.

The inadequacy of using human personality questionnaires for evaluating conversational agent personality is not only apparent in the lack of a dimension of perceived artificiality but also in some of the single personality questionnaire items, which were answered similarly across all three chatbots. Notably, some of these items for which participants did not perceive any difference state comparisons with humans, e.g. "is less active than other people". Albeit inventories are useful in collecting personality impressions, in particular because they allow for fast comparisons [67], our results suggest that the appropriateness of the items' phrasing should be reexamined for artificial agents.

In line with work on *speech-based* CA personalities [126], our findings indicate that the human Big Five model and corresponding questionnaires are likely not sufficient to paint a complete picture of *text-based* CA personalities. This insight further highlights the need for dedicated CA personality assessment tools. Our study has revealed insights into the advantage of open personality descriptions for examining the most salient personality traits in chatbots. On the other hand, dedicated inventories allow capturing the full range of perceived personality dimensions and foster a quick comparison of different agents. A possible next step towards this goal could be a continuation of our prior work on CA personality dimensions [126] by using our identified CA personality adjectives to adapt human personality questionnaires, which then have to be validated.

8 CONCLUSION

Whilst previous work on robots and voice assistants has emphasised the positive impact of certain personalities on the user experience, little has been known about personality perception of purely text-based chatbots. To address this gap, we contribute (1) a set of verbal cues derived from psycholinguistic literature to induce different levels of Extraversion implemented in a chatbot app and (2) a systematic empirical analysis of N=34 participants evaluating these chatbot personalities after interacting with them for four days each.

Our results show that participants perceived the extraverted and average chatbots as such, whereas verbal cues transferred from human behaviour were insufficient to create an introverted chatbot. Furthermore, our findings shed light on the salience of Agreeableness, Extraversion, and artificiality in chatbots for user preferences and engagement, thereby providing much-needed knowledge to CA designers regarding the personality traits on which to focus.

In a wider view, our study underlines the vision of imbuing CAs with different human-like personalities that can then be adapted to individual preferences to improve the overall interaction experience. However, future work in examining user characteristics is essential to better determine their preference for a particular CA personality.

REFERENCES

[1] David A. Aaker and Douglas M. Stayman. 1992. Implementing the concept of transformational advertising. *Psychology & Marketing* 9, 3 (1992), 237–253. https://doi.org/10.1002/mar.4220090306

[2] Rangina Ahmad, Dominik Siemon, Ulrich Gnewuch, and Susanne Robra-Bissantz. 2021. The Benefits and Caveats of Personality-Adaptive Conversational Agents in Mental Health Care. In *Twenty-Seventh Americas Conference on Information Systems*. Association for Information Systems, Atlanta, GA, USA, 10 pages. https://aisel.aisnet.org/amcis2021/sig_hci/sig_hci/21.

[3] Tagduda Ait Challal and Ouriel Grynszpan. 2018. What Gaze Tells Us About Personality. In *Proceedings of the 6th International Conference on Human-Agent Interaction* (Southampton, UK) *(HAI '18)*. Association for Computing Machinery, New York, NY, USA, 129–137. https://doi.org/10.1145/3284432.3284455

[4] David M. Almeida, Elaine Wethington, and Ronald C. Kessler. 2002. The Daily Inventory of Stressful Events: An Interview-Based Approach for Measuring Daily Stressors. *Assessment* 9, 1 (2002), 41–55. https://doi.org/10.1177/1073191102091006

[5] Tomás Alves, Bárbara Ramalho, Daniel Gonçalves, Sandra Gama, and Joana Henriques-Calado. 2020. Exploring How Personality Models Information Visualization Preferences. In *2020 IEEE Visualization Conference (VIS)*. IEEE, New York, NY, USA, 201–205. https://doi.org/10.1109/VIS47514.2020.00047

[6] Amir Aly and Adriana Tapus. 2013. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, New York, NY, USA, 325–332. https://doi.org/10.1109/HRI.2013.6483606

[7] Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. 2000. Integrating Models of Personality and Emotions into Lifelike Characters. In *Affective Interactions: Towards a New Generation of Computer Interfaces*, Ana Paiva (Ed.). Springer, Berlin, Heidelberg, Germany, 150–165. https://doi.org/10.1007/10720296_11

[8] Sean Andrist, Bilge Mutlu, and Adriana Tapus. 2015. Look Like Me: Matching Robot Personality via Gaze to Increase Motivation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3603–3612. https://doi.org/10.1145/2702123.2702592

[9] Michael Argyle. 1988. *Bodily communication*. Methuen & Co Ltd, London, UK.

[10] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies* 4, 3 (May 2009), 114–123. https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/.

[11] Samuel Bell, Clara Wood, and Advait Sarkar. 2019. Perceptions of Chatbots in Therapy. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313072

[12] Camiel J Beukeboom, Martin Tanis, and Ivar E Vermeulen. 2013. The language of extraversion: Extraverted people talk more abstractly, introverts are more concrete. *Journal of Language and Social Psychology* 32, 2 (2013), 191–201. https://doi.org/10.1177/0261927X12460844

[13] Timothy Bickmore and Justine Cassell. 2001. Relational Agents: A Model and Implementation of Building User Trust. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 396–403. https://doi.org/10.1145/365024.365304

[14] Timothy Bickmore and Justine Cassell. 2005. Social Dialogue with Embodied Conversational Agents. In *Advances in Natural Multimodal Dialogue Systems*, Jan C. J. van Kuppevelt, Laila Dybkjær, and Niels Ole Bernsen (Eds.). Springer, Dordrecht, Netherlands, 23–54. https://doi.org/10.1007/1-4020-3933-6_2

[15] Timothy W. Bickmore, Suzanne E. Mitchell, Brian W. Jack, Michael K. Paasche-Orlow, Laura M. Pfeifer, and Julie O'Donnell. 2010. Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers* 22, 4 (09 2010), 289–298. https://doi.org/10.1016/j.intcom.2009.12.001

[16] Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc., DE, USA.

[17] Ryan L Boyd and James W Pennebaker. 2017. Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences* 18 (2017), 63–68. https://doi.org/10.1016/j.cobeha.2017.07.017

[18] Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42, 9 (2010), 2355 – 2368. https://doi.org/10.1016/j.pragma.2009.12.012

[19] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland, UK) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 40, 11 pages. https://doi.org/10.1145/3290605.3300270

[20] Paul Bremner, Oya Celiktutan, and Hatice Gunes. 2016. Personality perception of robot avatar tele-operators. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE, New York, NY, USA, 141–148. https://doi.org/10.1109/HRI.2016.7451745

[21] John Brooke. 1996. SUS – A quick and dirty usability scale. *Usability evaluation in industry* 189 (1996), 4 – 7. Issue 194. https://doi.org/10.1201/9781498710411

[22] Donn Erwin Byrne. 1971. *The attraction paradigm*. Academic Press, Cambridge, MA, USA.

[23] Angelo Cafaro, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2016. First Impressions in Human–Agent Virtual Encounters. *ACM Trans. Comput.-Hum. Interact.* 23, 4, Article 24 (Aug. 2016), 40 pages. https://doi.org/10.1145/2940325

[24] Anne Campbell and J. Philippe Rushton. 1978. Bodily communication and personality. *British Journal of Social and Clinical Psychology* 17, 1 (1978), 31–36. https://doi.org/10.1111/j.2044-8260.1978.tb00893.x

[25] David W. Carment, C.G. Miles, and VB.. Cervin. 1965. Persuasiveness and persuasibility as related to intelligence and extraversion. *British Journal of Social and Clinical Psychology* 4, 1 (1965), 1–7. https://doi.org/10.1111/j.2044-8260.1965.tb00433.x

[26] Avshalom Caspi, Brent W. Roberts, L.A. Pervin, and O.P. John. 1990. Personality continuity and change across the life course. *Handbook of personality: Theory and research* 2 (1990), 300–326.

[27] Susana Castillo, Philipp Hahn, Katharina Legde, and Douglas W. Cunningham. 2018. Personality Analysis of Embodied Conversational Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (Sydney, NSW, Australia) *(IVA '18)*. Association for Computing Machinery, New York, NY, USA, 227–232. https://doi.org/10.1145/3267851.3267853

[28] Fabio Catania, Micol Spitale, and Franca Garzotto. 2021. Toward the Introduction of Google Assistant in Therapy for Children with Neurodevelopmental Disorders: An Exploratory Study. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 272, 7 pages. https://doi.org/10.1145/3411763.3451666

[29] William J. Chopik and Shinobu Kitayama. 2018. Personality change across the lifespan: Insights from a cross-cultural longitudinal study. *Journal of Personality* 86, 3 (2018), 508–521. https://doi.org/10.1111/jopy.12332

[30] Paul T. Costa and Robert R. McCrae. 1992. Four ways five factors are basic. *Personality and Individual Differences* 13, 6 (1992), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I

[31] Benjamin R. Cowan, Holly P. Branigan, Mateo Obregón, Enas Bugis, and Russell Beale. 2015. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human-computer dialogue. *International Journal of Human-Computer Studies* 83 (2015), 27 – 42. https://doi.org/10.1016/j.ijhcs.2015.05.008

[32] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 43, 12 pages. https://doi.org/10.1145/3098279.3098539

[33] Daniel Danner, Beatrice Rammstedt, Matthias Bluemke, Lisa Treiber, Sabrina Berres, Christopher Soto, and Oliver John. 2016. *Die deutsche Version des Big Five Inventory 2 (BFI-2)*. GESIS, Mannheim, Germany. https://doi.org/10.6102/zis247

[34] Boele de Raad. 2000. *The Big Five Personality Factors: The psycholexical approach to personality*. Hogrefe & Huber Publishers, Göttingen, Germany.

[35] Ian J. Deary. 2009. The trait approach to personality. In *The Cambridge handbook of personality psychology*, Philip J. Corr and Gerald Matthews (Eds.). Cambridge University Press, Cambridge, UK, Chapter 6, 89–109. https://doi.org/10.1017/CBO9780511596544.009

[36] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems* (Paris, France) *(AAMAS '14)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1061–1068. https://dl.acm.org/doi/10.5555/2615731.2617415.

[37] Jean-Marc Dewaele and Adrian Furnham. 1999. Extraversion: The unloved variable in applied linguistic research. *Language Learning* 49, 3 (1999), 509–544. https://doi.org/10.1111/0023-8333.00098

[38] Jean-Marc Dewaele and Adrian Furnham. 2000. Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences* 28, 2 (2000), 355–365. https://doi.org/10.1016/S0191-8869(99)00106-3

[39] Philip R Doyle, Leigh Clark, and Benjamin R. Cowan. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 244, 14 pages. https://doi.org/10.1145/3411764.3445206

[40] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 5, 12 pages. https://doi.org/10.1145/3338286.3340116

[41] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies* 132 (2019), 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

[42] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 4, 2 (06 Jun 2017), e19. https://doi.org/10.2196/mental.7785

[43] Adrian Furnham. 1990. Language and Personality. In *Handbook of language and social psychology*, William Peter Robinson and Howard Giles (Eds.). John Wiley & Sons, Chichester, UK, 73–95.

[44] Alastair J. Gill and Jon Oberlander. 2002. Taking care of the linguistic features of extraversion. In *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*. Routledge, UK, 363–368.

[45] Ulrich Gnewuch, Meng Yu, and Alexander Maedche. 2020. The Effect of Perceived Similarity in Dominance on Customer Self-Disclosure to Chatbots in Conversational Commerce. In *Proceedings of the 28th European Conference on Information Systems (ECIS)*. Association for Information Systems, Atlanta, GA, USA, 16 pages. https://aisel.aisnet.org/ecis2020_rp/53.

[46] Lewis R. Goldberg. 1981. Language and individual differences: The search for universals in personality lexicons. In *Review of Personality and Social Psychology*, L. Wheeler (Ed.). Vol. 2. Sage Publications, Beverly Hills, CA, USA, 141–166.

[47] Lewis R. Goldberg. 1982. From Ace to Zombie: Some explorations in the language of personality. In *Advances in Personality Assessment*, Charles D. Spielberger and James N. Butcher (Eds.). Vol. 1. Erlbaum, Hillsdale, NJ, USA, 203–234.

[48] Lewis R. Goldberg. 1990. An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59, 6 (1990), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

[49] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37, 6 (2003), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

[50] Sajanee Halko and Julie A. Kientz. 2010. Personality and Persuasive Technology: An Exploratory Study on Health-Promoting Mobile Applications. In *Persuasive Technology*, Thomas Ploug, Per Hasle, and Harri Oinas-Kukkonen (Eds.). Springer, Berlin, Heidelberg, Germany, 150–161. https://doi.org/10.1007/978-3-642-13226-1_16

[51] Xu Han, Michelle Zhou, Matthew J. Turner, and Tom Yeh. 2021. Designing Effective Interview Chatbots: Automatic Chatbot Profiling and Design Suggestion Generation for Chatbot Debugging. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 389, 15 pages. https://doi.org/10.1145/3411764.3445569

[52] Robert P. Hawkins, Matthew Kreuter, Kenneth Resnicow, Martin Fishbein, and Arie Dijkstra. 2008. Understanding tailoring in communicating about health. *Health education research* 23, 3 (2008), 454–466. https://doi.org/10.1093/her/cyn004

[53] Jacob B. Hirsh, Sonia K. Kang, and Galen V. Bodenhausen. 2012. Personalized Persuasion: Tailoring Persuasive Appeals to Recipients' Personality Traits. *Psychological Science* 23, 6 (2012), 578–581. https://doi.org/10.1177/0956797611436349

[54] Jacob B. Hirsh and Jordan B. Peterson. 2009. Personality and language use in self-narratives. *Journal of Research in Personality* 43, 3 (2009), 524–527. https://doi.org/10.1016/j.jrp.2009.01.006

[55] Thomas Holtgraves. 2011. Text messaging, personality, and the social context. *Journal of Research in Personality* 45, 1 (2011), 92–99. https://doi.org/10.1016/j.jrp.2010.11.015

[56] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: My Automated Conversation Coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 697–706. https://doi.org/10.1145/2493432.2493502

[57] Katherine Isbister and Clifford Nass. 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies* 53, 2 (2000), 251 – 267. https://doi.org/10.1006/ijhc.2000.0368

[58] Dipesh Kadariya, Revathy Venkataramanan, Hong Yung Yip, Maninder Kalra, Krishnaprasad Thirunarayanan, and Amit Sheth. 2019. KBot: Knowledge-enabled personalized chatbot for asthma self-management. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, New York, NY, USA, 138–143. https://doi.org/10.1109/smartcomp.2019.00043

[59] Ahmet Baki Kocabalil, Liliana Laranjo, and Enrico Coiera. 2018. Measuring user experience in conversational interfaces: a comparison of six questionnaires. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*. BCS Learning and Development Ltd., Belfast, UK, 1–12. https://doi.org/10.14236/ewic/HCI2018.21

[60] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C. Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The Personalization of Conversational Agents in Health Care: Systematic Review. *J Med Internet Res* 21, 11 (7 Nov 2019), e15360. https://doi.org/10.2196/15360

[61] Brigitte Krenn, Birgit Endrass, Felix Kistler, and Elisabeth André. 2014. Effects of Language Variety on Personality Perception in Embodied Conversational Agents. In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques. HCI 2014. Lecture Notes in Computer Science*, Masaaki Kurosu (Ed.), Vol. 8511. Springer International Publishing, New York, NY, USA, 429–439. https://doi.org/10.1007/978-3-319-07230-2_41

[62] Matthew W. Kreuter, David W. Farrell, Laura K. Brennan, and Laura R. Olevitch. 1999. *Tailoring health messages: Customizing communication with computer technology*. Routledge, UK.

[63] Rense Lange and Jason Rentfrow. 2007. Color and Personality: Strong's Interest Inventory and Cattell's 16PF. *North American Journal of Psychology* 9, 3 (2007), 423.

[64] Liliana Laranjo, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y. S. Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258. https://doi.org/10.1093/jamia/ocy072

[65] Kwan Min Lee and Clifford Nass. 2003. Designing Social Presence of Social Actors in Human Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 289–296. https://doi.org/10.1145/642611.642662

[66] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300932

[67] Kris Liu, Jackson Tolins, Jean E. Fox Tree, Michael Neff, and Marilyn A. Walker. 2016. Two Techniques for Assessing Virtual Agent Personality. *IEEE Transactions on Affective Computing* 7, 1 (2016), 94–105. https://doi.org/10.1109/TAFFC.2015.2435780

[68] Yun Liu, Lu Wang, William R. Kearns, Linda Wagner, John Raiti, Yuntao Wang, and Weichao Yuwen. 2021. Integrating a Voice User Interface into a Virtual Therapy Platform. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 362, 6 pages. https://doi.org/10.1145/3411763.3451595

[69] Gale M. Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100. https://doi.org/10.1016/j.chb.2014.04.043

[70] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[71] David D. Luxton. 2014. Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine* 62, 1 (2014), 1–10. https://doi.org/10.1016/j.artmed.2014.06.004

[72] David D. Luxton. 2020. Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization* 98, 4 (2020), 285 – 287.

[73] Davide Marengo, Fabrizia Giannotta, and Michele Settanni. 2017. Assessing personality using emoji: An exploratory study. *Personality and Individual Differences* 112 (2017), 74 – 78. https://doi.org/10.1016/j.paid.2017.02.037

[74] Robert R. McCrae. 2009. The Five-Factor Model of personality traits: consensus and controversy. In *The Cambridge Handbook of Personality Psychology*, Philip J. Corr and Gerald Matthews (Eds.). Cambridge University Press, Cambridge, UK, 148–161. https://doi.org/10.1017/CBO9780511596544.012

[75] Robert R. McCrae and Paul T. Costa. 2008. A five-factor theory of personality. In *Handbook of Personality: Theory and Research*, O.P. John, R.W. Robins, and L.A. Pervin (Eds.). Vol. 3. The Guilford Press, New York, NY, USA, 159–181.

[76] Robert R McCrae and Paul T Costa Jr. 1997. Personality trait structure as a human universal. *American psychologist* 52, 5 (1997), 509 – 516.

[77] Robert R. McCrae and Oliver P. John. 1992. An introduction to the five-factor model and its applications. *Journal of Personality* 60, 2 (1992), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

[78] Richard McElreath. 2016. *Statistical rethinking: A Bayesian course with examples in R and Stan.* Chapman and Hall/CRC, London, UK. https://doi.org/10.1201/9781315372495

[79] Margaret McRorie, Ian Sneddon, Gary McKeown, Elisabetta Bevacqua, Etienne de Sevin, and Catherine Pelachaud. 2012. Evaluation of Four Designed Virtual Agent Personalities. *IEEE Transactions on Affective Computing* 3, 3 (2012), 311–322. https://doi.org/10.1109/T-AFFC.2011.38

[80] Matthias R. Mehl, Samuel D. Gosling, and James W. Pennebaker. 2006. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology* 90, 5 (2006), 862–877. https://doi.org/10.1037/0022-3514.90.5.862

[81] Clifford Nass and Scott Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship.* MIT press, Cambridge, MA, USA.

[82] Clifford Nass and Kwan Min Lee. 2001. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied* 7, 3 (2001), 171. https://doi.org/10.1037/1076-898X.7.3.171

[83] Clifford Nass, Youngme Moon, B.J. Fogg, Byron Reeves, and D.Christopher Dryer. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies* 43, 2 (1995), 223–239. https://doi.org/10.1006/ijhc.1995.1042

[84] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703

[85] Michael Neff, Nicholas Toothman, Robeson Bowmani, Jean E. Fox Tree, and Marilyn A. Walker. 2011. Don't Scratch! Self-adaptors Reflect Emotional Stability. In *Intelligent Virtual Agents. IVA 2011. Lecture Notes in Computer Science*, Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella, and Kristinn R. Thórisson (Eds.), Vol. 6895. Springer, Berlin, Heidelberg, 398–411. https://doi.org/10.1007/978-3-642-23974-8_43

[86] Kate G. Niederhoffer and James W. Pennebaker. 2002. Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology* 21, 4 (2002), 337–360. https://doi.org/10.1177/026192702237953

[87] Warren T. Norman. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology* 66, 6 (1963), 574–583. https://doi.org/10.1037/h0040291

[88] Jon Oberlander and Alastair J. Gill. 2004. Individual differences and implicit language: personality, parts-of-speech and pervasiveness. *Proceedings of the Annual Meeting of the Cognitive Science Society* 26 (2004), 1035–1040. https://escholarship.org/uc/item/94c490mq.

[89] Yoshito Ogawa, Kouki Miyazawa, and Hideaki Kikuchi. 2014. Assigning a Personality to a Spoken Dialogue Agent through Self-Disclosure of Behavior. In *Proceedings of the Second International Conference on Human-Agent Interaction* (Tsukuba, Japan) *(HAI '14)*. Association for Computing Machinery, New York, NY, USA, 331–337. https://doi.org/10.1145/2658861.2658884

[90] World Health Organisation. 2021. Mental health preparedness and response for the COVID-19 pandemic: Report by the Director-General. https://apps.who.int/gb/ebwha/pdf_files/EB148/B148_20-en.pdf, last accessed April 10, 2022.

[91] World Health Organization. 2019. *The WHO special initiative for mental health (2019-2023): universal health coverage for mental health.* Technical documents. World Health Organization. 7 pages.

[92] Nirmita Panchal, Rabah Kamal, Kendal Orgera, Cynthia Cox, Rachel Garfield, and Liz Hamel. 2020. The implications of COVID-19 for mental health and substance use. https://abtcounseling.com/wp-content/uploads/2020/09/The-Implications-of-COVID-19-for-Mental-Health-and-Substance-Use-_-KFF.pdf, last accessed April 10, 2022.

[93] M. Patterson and D.S. Holmes. 1966. Social Interaction Correlates of the MMPI Extraversion Introversion Scale. *American Psychologist* 21 (1966), 724–25.

[94] James W. Pennebaker and Laura A. King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* 77, 6 (1999), 1296–1312. https://doi.org/10.1037/0022-3514.77.6.1296

[95] Marta Garcia Perez, Sarita Saffon Lopez, and Hector Donis. 2018. Voice Activated Virtual Assistants Personality Perceptions and Desires: Comparing Personality Evaluation Frameworks. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference* (Belfast, UK) *(HCI '18)*. BCS Learning & Development Ltd., Swindon, GBR, Article 40, 10 pages. https://doi.org/10.14236/ewic/HCI2018.40

[96] Laura M. Pfeifer and Timothy Bickmore. 2009. Should Agents Speak Like, um, Humans? The Use of Conversational Fillers by Virtual Agents. In *Intelligent Virtual Agents*, Zsófia Ruttkay, Michael Kipp, Anton Nijholt, and Hannes Högni Vilhjálmsson (Eds.). Springer, Berlin, Heidelberg, Germany, 460–466. https:

//doi.org/10.1007/978-3-642-04380-2_50

[97] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[98] Alisha Pradhan and Amanda Lazar. 2021. Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 12, 4 pages. https://doi.org/10.1145/3469595.3469607

[99] Byron Reeves and Clifford Ivar Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press, Cambridge, UK.

[100] Stuart Reeves. 2019. Conversation Considered Harmful?. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) *(CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 10, 3 pages. https://doi.org/10.1145/3342775.3342796

[101] F. Dan Richard, Charles F. Bond Jr, and Juli J. Stokes-Zoota. 2003. One hundred years of social psychology quantitatively described. *Review of general psychology* 7, 4 (2003), 331–363.

[102] Lionel Robert, Rasha Alahmad, Connor Esterwood, Sangmi Kim, Sangseok You, and Qiaoning Zhang. 2020. *A Review of Personality in Human–Robot Interactions.* Now Publishers, Boston, MA, USA. https://doi.org/10.2139/ssrn.3528496 arXiv:2001.11777

[103] Elayne Ruane, Sinead Farrell, and Anthony Ventresque. 2021. User Perception of Text-Based Chatbot Personality. In *Chatbot Research and Design.* Springer International Publishing, Cham, Switzerland, 32–47. https://doi.org/10.1007/978-3-030-68288-0_3

[104] Peter A. M. Ruijten. 2021. The similarity-attraction paradigm in persuasive technology: effects of system and user personality on evaluations and persuasiveness of an interactive system. *Behaviour & Information Technology* 40, 8 (2021), 734–746. https://doi.org/10.1080/0144929X.2020.1723701

[105] D. R. Rutter, Ian E. Morley, and Jane C. Graham. 1972. Visual interaction in a group of introverts and extraverts. *European Journal of Social Psychology* 2, 4 (1972), 371–384. https://doi.org/10.1002/ejsp.2420020403

[106] Klaus Rainer Scherer. 1979. Personality markers in speech. In *Social Markers in Speech*, Klaus Rainer Scherer and Howard Giles (Eds.). Cambridge University Press, Cambridge, UK.

[107] Klaus Rainer Scherer. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 1 (2003), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

[108] Michael Schmitz, Antonio Krüger, and Sarah Schmidt. 2007. Modelling Personality in Voices of Talking Products through Prosodic Parameters. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (Honolulu, Hawaii, USA) *(IUI '07)*. Association for Computing Machinery, New York, NY, USA, 313–316. https://doi.org/10.1145/1216295.1216355

[109] Hanna Schneider, Katrin Schauer, Clemens Stachl, and Andreas Butz. 2017. Your Data, Your Vis: Personalizing Personal Data Visualizations. In *Human-Computer Interaction – INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer International Publishing, Cham, 374–392. https://doi.org/10.1007/978-3-319-67687-6_25

[110] Jessica Schroeder, Chelsey Wilkes, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M. Linehan. 2018. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3173574.3173972

[111] Katie Seaborn and Jacqueline Urakami. 2021. Measuring Voice UX Quantitatively: A Rapid Review. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, Article 416, 8 pages. https://doi.org/10.1145/3411763.3451712

[112] Michael Shumanov and Lester Johnson. 2021. Making conversations with chatbots more personalized. *Computers in Human Behavior* 117 (2021), 106627. https://doi.org/10.1016/j.chb.2020.106627

[113] Christopher J. Soto and Oliver P. John. 2017. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology* 113, 1 (2017), 117 – 143. https://doi.org/10.1037/pspp0000096

[114] Christopher J. Soto and Oliver P. John. 2017. Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality* 68 (2017), 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

[115] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis. 2020. User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *J Med Internet Res* 22, 3 (6 Mar 2020), e16235. https://doi.org/10.2196/16235

[116] Adriana Tapus, Cristian Ţăpuş, and Maja J. Matarić. 2008. User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics* 1, 2 (13 Feb 2008), 169. https://doi.org/10.1007/s11370-008-0017-4

[117] Jürgen Trouvain, Sarah Schmidt, Marc Schröder, Michael Schmitz, and William J. Barry. 2006. Modelling personality features by changing prosody in synthetic speech. In *Proceedings of the 3rd International Conference on Speech Prosody.* TUDpress, Dresden, Germany, 4 pages. https://doi.org/10.22028/D291-25920

[118] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Comput. Surv.* 50, 6, Article 93 (Dec. 2017), 40 pages. https://doi.org/10.1145/3123988

[119] Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 273–291. https://doi.org/10.1109/TAFFC.2014.2330816

[120] Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schroeder. 2012. Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Transactions on Affective Computing* 3, 1 (2012), 69–87. https://doi.org/10.1109/T-AFFC.2011.27

[121] James Vlahos. 2019. *Talk to Me: How Voice Computing Will Transform the Way We Live, Work, and Think.* Houghton Mifflin Harcourt, Boston, MA, USA.

[122] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 254, 15 pages. https://doi.org/10.1145/3411764.3445536

[123] Sarah Theres Völkel, Daniel Buschek, Jelena Pranjic, and Heinrich Hussmann. 2019. Understanding Emoji Interpretation through User Personality and Message Context. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 3, 12 pages. https://doi.org/10.1145/3338286.3340114

[124] Sarah Theres Völkel and Lale Kaya. 2021. Examining User Preference for Agreeableness in Chatbots. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 38, 6 pages. https://doi.org/10.1145/3469595.3469633

[125] Sarah Theres Völkel, Samantha Meindl, and Heinrich Hussmann. 2021. Manipulating and Evaluating Levels of Personality Perceptions of Voice Assistants through Enactment-Based Dialogue Design. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. https://doi.org/10.1145/3469595.3469605

[126] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Bühner, and Heinrich Hussmann. 2020. Developing a Personality Model for Speech-Based Conversational Agents Using the Psycholexical Approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376210

[127] Joseph B Walther. 2011. Theories of computer-mediated communication and interpersonal relations. In *The handbook of interpersonal communication.* Sage Publications Ltd., Thousand Oaks, CA, USA, Chapter 4, 443–479.

[128] Yingying Wang, Jean E. Fox Tree, Marilyn Walker, and Michael Neff. 2016. Assessing the Impact of Hand Motion on Virtual Character Personality. *ACM Trans. Appl. Percept.* 13, 2, Article 9 (March 2016), 23 pages. https://doi.org/10.1145/2874357

[129] Joseph Weizenbaum. 1976. *Computer power and human reason: From judgment to calculation.* WH Freeman & Co, New York, NY, USA.

[130] Ziang Xiao, Michelle X. Zhou, Q. Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-Ended Questions. *ACM Trans. Comput.-Hum. Interact.* 27, 3, Article 15 (June 2020), 37 pages. https://doi.org/10.1145/3381804

[131] Tal Yarkoni. 2010. Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality* 44, 3 (2010), 363 – 373. https://doi.org/10.1016/j.jrp.2010.04.001

[132] Renwen Zhang, Kathryn E. Ringland, Melina Paan, David C. Mohr, and Madhu Reddy. 2021. Designing for Emotional Well-Being: Integrating Persuasion and Customization into Mental Health Technologies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, Article 542, 13 pages. https://doi.org/10.1145/3411764.3445771

[133] Michelle X. Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting Virtual Agents: The Effect of Personality. *ACM Trans. Interact. Intell. Syst.* 9, 2-3, Article 10 (March 2019), 36 pages. https://doi.org/10.1145/3232077

A APPENDIX

Table 1: Participants' open personality descriptors for the three chatbots were mapped to the Big Five dimensions (high and low levels) and their subclusters. Please note that descriptors were reproduced in participants' own words but were partly assigned to the clusters based on synonyms. The counts refer to the number of times descriptors for this cluster were mentioned, independent of whether they were uttered by the same participant.

			Introverted Chatbot		Average Chatbot		Extraverted Chatbot	
	Level	**Cluster**	**Descriptors**	**#**	**Descriptors**	**#**	**Descriptors**	**#**
Openness	high	Creativity		0		0	modern	1
		Curiosity	inquisitive, interested	7	nosy, interested	2	inquisitive, enquiring, interested	6
		Insight	perceptive, knowledgeable	2		0	perceptive	1
		Intelligence	intelligent	1	clever	1	clever	1
Conscientiousness	high	Dignity	formal	6	formal	3	formal	1
		Persistence	persistent, thorough, in-depth	3	persistent	2	thorough	1
		Predictability	consistent	1		0	consistent	1
		Efficiency	professional, concise, straight to the point	7	straight to the point, efficient, professional, succinct, effective	12		0
		Organisation		0	organised	1		0
		Precision	factual, focused, accurate	3	accurate, focused	3		0
		Punctuality	speedy	1	quick	1		0
		Decisiveness	strong	1		0		0
		Dependability	reliable	1		0		0
		Logic	logical, analytical	2		0		0
	low	Disorganisation		0		0	ditzy	1
		Negligence		0	uncaring	1	unprofessional	1
Extraversion	high	Assertion		0	assertive	3	assertive, affirmative	2
		Candour	direct	1	open, straightfoward, direct	5	open	2
		Expressiveness		0		0	communicative	1
		Gregariousness		0		0	sociable, outgoing	2
		Optimism	upbeat, positive	2		0	happy, upbeat, cheerful	7
		Spirit		0		0	enthusiastic, overenthusiastic, bubbly	3
		Spontaneity		0		0	spontaneous	1
		Talkativeness		0	talkative	1	chatty	2
		Humour		0	funny	1		0
		Bossiness	demanding	1		0		0
	low	Reserve	detached, neutral	4	neutral	2		0
		Silence		0	calm	2		0
		Inhibition	measured	1		0		0
Agreeableness	high	Amiability	friendly, lovely	6	friendly, approachable, inoffensive	8	friendly, over-friendly, lovely	22
		Cooperation	helpful, sensible	3	helpful, supportive	3	helpful, personable	8
		Empathy	understanding, thoughtful, sympathetic, kind	5	nice, kind, thoughtful	8	trusting, kind	2
		Generosity	polite	1		0	polite	2
		Morality	honest	1	fair	1	genuine	1
		Naturalness		0		0	informal, casual, relaxed, over-familiar	6
		Warmth	sensitive, caring	3		0	caring	2
		Courteous	attentive	1		0		0
	low	Irritability		0		0	kooky	1
		Callousness	impersonal, cold	2	cold, impersonal	4		0
		Surliness	blunt	1	blunt	1		0
		Unfriendliness	unfriendly	1	unfriendly	2		0
Neur.	high	Intrusiveness		0	invasive	1		0
	low	Placidity	matter of fact	3	matter of fact	2	emotionless	1
Artif.	high	Thingness	unnatural, automated, robotic	4	robotic, automated, clinical, no personality, generic, unnatural	10	unnatural, robotic, automated	3
	low	Humanness	human-like	1	relatable	1		0
Usability	high	Usable	easy to understand, informative, responsive, reactive	4	responsive, intuitive	3	informative	1
	low	Unusable	dysfunctional	1		0	repetitive	1
Others			un-engaging	1	black and white, inclusive, inviting, not understand emotional side of how I was feeling, gets one thinking, objective, reassuring	7	engaging, flirty, trying to be cool, gave more in-depth information about herself	4