

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/348965237>

User Perception of Text-Based Chatbot Personality

Chapter · February 2021

DOI: 10.1007/978-3-030-68288-0_3

CITATIONS

24

READS

3,462

3 authors, including:



[Elayne Ruane](#)

University College Dublin

9 PUBLICATIONS 196 CITATIONS

[SEE PROFILE](#)



[Sinead Farrell](#)

University College Dublin

1 PUBLICATION 24 CITATIONS

[SEE PROFILE](#)

User Perception of Text-Based Chatbot Personality

Elayne Ruane^{1,2}, Sinead Farrell¹, and Anthony Ventresque^{1,2}

¹ School of Computer Science, University College Dublin, Ireland

² LERO - The Irish Software Research Centre

{elayne.ruane, sinead.farrell1}@ucdconnect.ie

Abstract. This work explores the effect of chatbot personality on user experience and investigates how users perceive agent personality when conveyed through text. Building on previous work in the field of human-computer interaction on designing chatbot personality, we investigate whether users in a low-stakes conversation have a preference for a specific personality type when the agent does not use voice, is not visually represented, and does not provide identity cues such as gender. We developed two chatbots that interact with users in a multi-turn conversation and designed them to have distinct personalities along two axes of the Five Factor Model (extraversion and agreeableness). We conducted a user study to evaluate user engagement, user perception of the agents, and the effect of user personality on user experience.

Keywords: Conversational Agent · Chatbot · Personality · User Experience · HCI

1 Introduction

Chatbots are designed to mimic a uniquely human activity: conversation. Human-human communication is full of complexity and nuance. When we interact with one another, our personalities inform how we build connections and form relationships [16]. Our behaviours, and how we interpret the behaviour of others, is heavily influenced by our personality [1] [17]. It follows that when a human user is interacting with, and forming their perception of, a chatbot that is imitating human behaviour, they may infer personality traits from its language and response style, as well as from other anthropomorphic cues such as visual representations of the chatbot, including non-verbal behaviours like animated facial expressions and gestures. Additionally, agents with voice capabilities have other features including tone, pitch, and cadence that may influence user perception by cuing gender, age, and other perceived identity markers.

A person’s personality has the power to sway the direction of a conversation [16]. Correspondingly, in commercial contexts such as customer service, a human representative’s personality can be the key to ensuring a satisfying user experience [27]. The literature, as described in Section 2, suggests the same is true for chatbots, in both commercial and other scenarios. However, designing

chatbot personality that maximises user satisfaction and provides an engaging user experience and, thus, a better service, is a non-trivial task. Many interesting questions remain around how to choose an appropriate personality for a chatbot and how to design dialogue that reliably conveys this simulated personality.

While much of the literature focuses on multi-modal agents that have several avenues to leverage in expressing personality, we are interested in text-based agents that do not have a visual or audio representation. Text-based agents are pervasive. For example, the popular entertainment chatbot and five-time winner of the Loebner Prize Turing Test, Mitsuku³, is a text-based chatbot (albeit with an avatar but without voice capabilities), as are many recommendation agents⁴, information retrieval chatbots (e.g. news⁵ and weather⁶), therapy chatbots⁷, and customer service and FAQ chatbots embedded in company websites. As such, personality design and user perception of personality in text-based agents warrants investigation. To this end, we have defined two research questions; RQ1) Can personality be reliably simulated by a chatbot via text such that the user perceives personality or personality traits as intended? and RQ2) Does the perceived personality of a text-based chatbot affect user experience?

To address these questions, we have developed two text-based chatbots (i.e. without voice capabilities and not represented with a visual avatar) with distinct personalities and conducted a user study to evaluate whether users perceive the personality traits expressed through dialogue design as intended and whether they exhibit a preference for one personality over another. The rest of this paper is structured as follows: in Section 2 we discuss related work, in Section 3 we outline the design of the chatbots used in this study, in Section 4 we detail our methodology and experiment design, we provide the results and a discussion of those results in Section 5 and 6 respectively, in Section 7 we discuss the limitations of this study, and lastly we conclude the work in Section 8.

2 Related Work

In this section, we discuss previous work that aims to understand how users perceive and interact with chatbots, specifically studies that investigate how chatbot personality design affects user experience, how users perceive agent personality, and how agent personality can be conveyed through text.

2.1 Five Factor Model Studies

Many studies of chatbot personality use the well-established Five Factor Model (FFM), also referred to as the Big Five Trait Taxonomy [9], to model both agent and user personality. The model consists of five characteristics: openness

³ Mitsuku: <https://www.pandorabots.com/mitsuku/>

⁴ Beauty recommendation chatbots: <https://beauty.bot/directory/makeup/>

⁵ CNN Chatbot: <https://marutitech.com/news-made-personal-with-chatbots/>

⁶ Weather chatbots: <https://chatbottle.co/bots/messenger/weather>

⁷ Woebot: <https://woebothealth.com/>

to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Each trait is a continuum describing a dimension of the most common traits perceived in individuals. Trait-based models of personality such as FFM are widely used in both Psychology and in Affective Computing. Such models are useful in evaluating individual differences [4] and thus lend themselves to design of agent personality and investigations of personality effects on user experience.

Hanna and Richards (2015) [6] investigated the effect of agent personality on team work, specifically the development of a shared mental model between human users and a virtual agent, while completing a collaborative game task. Two dimensions of FFM, extraversion and agreeableness, were expressed by the agent using both verbal and non-verbal cues. The authors found participants were able to identify the personality traits as intended and that an agent designed with explicit personality traits is likely to improve team performance.

Kang et al. (2008) [10] found participant personality traits (modelled by FFM) affect their sense of rapport with, and their perception of, a virtual agent, regardless of the agent’s personality design. This is supported by later work in which Von der Pütten et al. (2010) [22] investigated how participant personality, gender, and age affect both their behaviour when interacting with a virtual agent and also their evaluation of that agent. They found gender and age did not affect the evaluation but some personality traits were predictive, including agreeableness and extraversion where agreeableness had a positive impact on how participants perceived the interaction and extraversion impacted participant’s verbal behaviour, in particular, the number of words they used. The agent in this study uses both verbal and non-verbal cues but only the non-verbal cues were varied. It should be noted that while gender was found not to be a predictive factor, the agent was coded female, including using a female voice, and research has shown users treat female- and male-coded systems differently [25].

Other studies have investigated whether a “match” in user-agent personality improves user experience. Isbister and Nass (2000) [8] studied how perceived extraversion/introversion of an agent affected user experience in a low-stakes discussion task. They found users prefer consistency across both verbal and non-verbal personality cues and prefer a personality complementary to theirs, rather than entirely similar. Similarly, Liew and Tan (2016) [12] developed two pedagogical virtual agents, one introverted and the other extroverted, also expressed using both verbal and non-verbal cues. The results of the study support the complementary-attraction principle such that learners’ experience was improved when the agent’s personality complemented their own. These studies that include visual non-verbal cues draw conclusions that support the complementary-attraction principle, unlike previous work [11] [18] that shows users prefer agents that exhibit similar personality traits (similarity-attraction principle) when communicated through voice and text only.

Smestad and Volden (2018) [26] investigated the effect of a match in personality between a chatbot and the user where the chatbot is representing a brand. The authors used the FFM model and created two chatbots, one with an “agreeable” personality and the other with a “conscientious” personality. The

chatbots were varied to different degrees across the five dichotomies; the neuroticism trait was excluded for one chatbot and both chatbots appear to have high conscientiousness. The main difference between the two chatbots is one is high in agreeableness and extraversion and the other is low on those traits. This study did not use non-verbal cues but varied across lexical features and also used voice, leveraging *tone* of voice specifically. The authors acknowledge previous work [25] that has shown female-coded agents are more likely to be stereotyped and receive abusive messages and although gender was not included as a factor in the study, both chatbots were represented with a human avatar and coded as female. The authors found the agreeable personality had a more positive effect on user experience with the particular brand and user group involved in this study.

2.2 Expressing Personality Through Text

Many studies in this area, including those discussed above, focus on voice-based agents and leverage visual cues such as animated facial expressions or body language to express personality. However, many chatbots in the wild do not use these cues, instead relying largely or solely on text to convey personality. Neff et al. (2010) [20] identified verbal and non-verbal cues that can be used to demonstrate extraversion in a chatbot. Drawing on the Psychology literature, the authors detail the linguistic parameters of their language generation model used to display extraversion. These include high verbosity, content polarity, and acknowledgements, along with low negation, filled pauses, and softener hedges.

Roffo et al. (2014) [23] explored identification of personality in textual human-human conversation via three stylometric features. We apply these features to human-agent conversation in both the expression of personality on the part of the agent, and as features in our analysis of user behaviour. *Lexical* features such as the number of words or characters used per turn may be a sign of user engagement but from the user’s perspective, these features may convey personality traits of the chatbot and may also be linked to how informative or effective the agent is. *Syntactic* features such as the use of emoticons or expressive punctuation can be used to convey emotion or sentiment. Lastly, *turn-taking* features include turn duration and answer time, which will vary markedly for the user and the chatbot. In this study, we use these features to inform the dialogue design and to analyse the conversation logs from the study.

2.3 Interaction Questionnaire Design

Liu et al. (2015) [13] compared two types of questionnaires for measuring user perception of agent personality; open-ended questions and Likert-scale personality inventories (such as the Big Five Inventory). The personality traits examined in this study were extraversion/introversion and neuroticism. The authors found both question styles yield different yet complementary results. The open-ended questions do not prime participants and thus give insight into aspects of personality or agent design that resonated most with the participant. Personality

inventories can be useful for eliciting opinion on traits that may have been observed but subsequently forgotten or were not otherwise verbalized. However, such inventories can also prompt the participant to think of the agent in a way they had not previously. Based on this work, and as we are conducting a within-subject study, we elected to use open-ended questions and did not ask users to fill out an agent personality inventory.

Luger et al. (2016) [14] interviewed conversational agent users to understand their experience with agents like Siri, Google Now, Cortana, and Alexa. The authors found higher user expectations of agent capabilities can lead to lower user satisfaction. They discuss why expectations may go unmet such as a gap in the user’s understanding of agent capabilities due to a lack of technical knowledge or experience. Based on these findings, we include two questions in our pre-interaction demographic questionnaire that ask users to (i) describe their understanding of a chatbot and (ii) describe their frequency of use. This allows us to understand their mental model and evaluate if their previous experience affects their perception of the interaction.

3 Design and Development of Chatbots with Personality

We designed and implemented our chatbots using the Microsoft Bot Framework⁸ and its NLP service, LUIS⁹. We carefully considered the impact of the application domain when selecting the topic of conversation for this study. We did not want the conversation to be high-stakes, commercially driven, or to focus on a strategic task. We wanted participants to have a truly conversational experience (as opposed to conversational search or a button-based interaction). It was not feasible for us to build an open-domain agent, thus we designed the bots to have a multi-turn conversation about a specific topic familiar to the participants; third-level education, or, more specifically, computer science courses and university campus experiences.

As we are investigating the effect of personality on user experience, we endeavoured to mitigate any other persona-related effect. Previous work [27] has shown the use of an avatar can both positively or negatively impact how users perceive a chatbot, even before they interact with it. Silvervarg et al. (2012) [25] have shown how female-coded agents are treated more poorly than male-coded agents. As a result, we have not provided visual representation of the agents, have not used identity-specific language, and have given the chatbots androgynous names¹⁰.

In chatbot personality design, there are limits to the expression of some personality traits through text dialogue [7], even more so when designing dialogue

⁸ Microsoft Bot Framework: <https://dev.botframework.com>

⁹ <https://www.luis.ai/>

¹⁰ The bots were called Makoto and Nasoto. Makoto is an ungendered Japanese name, and Nasoto is a non-word. These names are culturally distant from the cultural background of our sample and thus unlikely to be readily associated with a specific gender, age-group, or other identity.

for a domain not usually associated with emotive language. As such, some traits in the FMM may be more difficult to express than others and some more appropriate than others. For example, high neuroticism may be inappropriate for an agent designed to assist the user with a routine task. With this in mind, and in context of the related work discussed in Section 2.1, we focus on two dimensions of the FFM; extraversion and agreeableness.

Extraversion can be broken down into five distinct components: activity level, dominance, sociability, expressiveness, and positive emotionality [9]. Agreeableness includes traits relating to trust, altruism, compliance, modesty, and tender mindedness [9]. Tables 1 and 2 show language cues that distinguish extroverts from introverts, and highly agreeable people from those who are less agreeable. These cues were compiled from work in the Linguistics and Psychology literature [2, 3, 5, 19, 21]. Although a 2x2 factorial design such that participants interact with agent personalities that vary across all combinations of extraversion and agreeableness may be of interest, we decided to create two chatbots with personalities that combine high extraversion with high agreeableness (and vice versa) due to their complementary linguistic presentations and previous work ([15] [14]) that suggests users respond to *distinct* agent personalities.

The applicable speech patterns and response styles were applied to the conversation flow for each chatbot to generate responses that demonstrate a consistent, distinct personality. Chatbot A was designed to exhibit high extraversion by demonstrating (i) high energy through punctuation including exclamation points, (ii) a talkative nature through verbosity of phrasing, and (iii) sharing information by asking questions of the user. High agreeableness is shown through complementary language and positive reinforcement. Chatbot B was designed to exhibit a contrasting personality with low extraversion and low agreeableness. As such, the dialogue is designed to demonstrate low energy, passiveness, and overall show less interest in participating in chitchat than chatbot A. Chatbot B uses a direct style of communication with less interest in the user as an individual; the questions posed are more factual in nature, rather than personal to the user.

By way of example, when the user tells the chatbot how many courses they are taking in the current semester, Chatbot A may respond “Wow <num> modules! Which one would you say is your favourite?” (See Figure 1 for other response examples). This response is (i) relatively informal (ii) uses an exclamation mark (iii) contains no negations and (iv) uses cheerful, positive language. In comparison, Chatbot B may respond “Which module is your favourite? Mine would be secure software engineering” (See Figure 2 for other response examples). This response is (i) self-focused (ii) more direct, and (iii) the language is less positive. While the response text differs, both chatbots follow the same conversation flow and can discuss the same scope of topics about university life including course modules, exams, the campus, and extracurricular activities. In addition, the chatbots can discuss the Covid-19 pandemic and how it has impacted the previously listed topics.

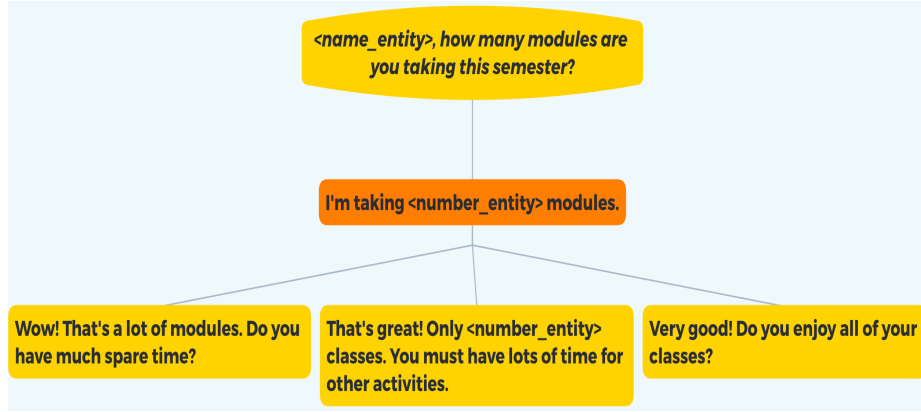


Fig. 1. An example conversation flow snippet for Chatbot A (high extraversion and high agreeableness).

4 Methodology

The experiment was designed to provide insight into user perception of the chatbots including their attitudes towards the agents and their behaviour while interacting with them. The participants were final year undergraduate computer science students in University College Dublin selected using convenience sampling via both email mailing lists and shared forum. A total of 22 people signed up to partake in the study, 5 of whom failed to complete all required components and so were excluded from the analysis (thus $n=17$). One participant fell in the 35-39 years age range while the rest were ages 18-24. The sample was comprised of 12 males and 5 females. As the scope of the conversation is the participants' college experience, they already possessed the required domain knowledge to interact with the chatbot and were familiar with the knowledge base of the agent. The study is a within-subject study such that each participant interacts with both chatbots.

Task 1: Pre-interaction Questionnaire. To begin, participants fill out a pre-interaction questionnaire that gathers participant demographic data including age-range and gender. Participants are also asked to detail their previous experience with and understanding of chatbots, including their frequency of use. Lastly, users fill out a personality inventory. We considered several personality questionnaires proposed in the literature including Eysenck's EPQ-R, the NEO-PI-R model, and the Big-Five Inventory (BFI). We decided to use the BFI as it is available freely for use in research, is based on the same model used to design the personalities of the chatbots, and has been used in similar work (see Section 2). The BFI uses a Likert-scale questionnaire through which the participants self-report their personality traits. While there are limitations around self-reporting, it is an accepted practice for subjective measures. The data col-

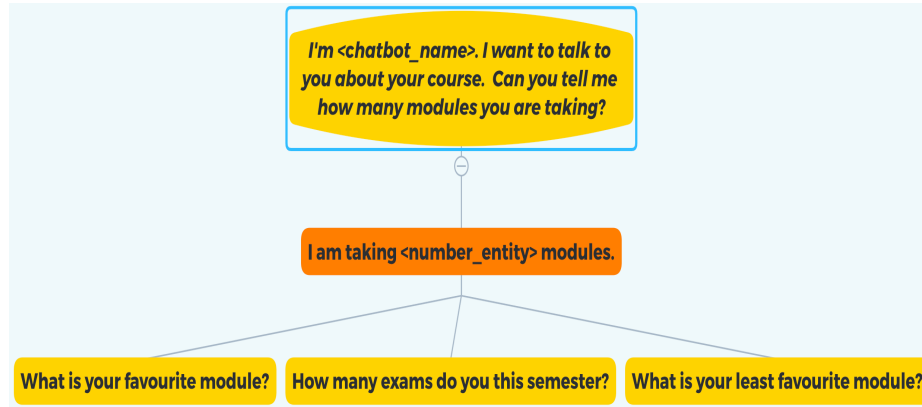


Fig. 2. An example conversation flow snippet for Chatbot B (low extraversion and low agreeableness).

lected from the pre-interaction questionnaire was analyzed to determine whether participant demographics, previous experience, or personality has a modulating effect on participant perception of, or behaviour with, each chatbot.

Task 2: Interaction with the Chatbot. For each chatbot the participants were instructed to click on the link associated with the chatbot interface. Once the interface had loaded they were directed to an input box where they began the conversation by typing and sending any message to the chatbot. They were advised that the conversation would be directed primarily by the chatbot asking questions about their university experience. Below are the instructions given to participants:

When you click the link below a white screen will appear, on the bottom of this screen there will be an input box that says "Type your message". When you are ready to begin your conversation type anything into the input box and click enter to send the message. Converse with the chatbot by answering its questions about your university experience. When you are finished conversing with the chatbot close the window.

Task 3: Post-Interaction Questionnaire. After the chatbot interaction, participants complete a post-interaction questionnaire about their experience that asks them (i) to describe the chatbot in an open-text field, (ii) whether they enjoyed the interaction, and (iii) to rate the chatbot across three dimensions (knowledge, quality of conversation, and attitude/personality) on a Likert scale. The participants also have the option to qualitatively expand on each rating in open-text fields. The results from the post-interaction surveys were analysed to determine participant attitudes to each chatbot and the descriptions were analysed for language describing demographic features such as age, gender, and cultural background that users may have ascribed to the chatbot(s).

Extraversion	Introversion
self-references	few self-references
informal	formal
pleasure talk	problem talk
agreement	listens
few tentative words	many tentative words
talks more about social topics	self-focused
many verbs, adverbs, pronouns	many nouns, adjectives, prepositions
few words per sentence	many words
few negations, articles	many negations, article
many social words	few social words
many positive emotion words	many negative emotion words

Table 1. Extraversion vs Introversion Language Cues

Agreeableness	Disagreeableness
prompting	negations
listens more	swearing
longer words	words relating to anger
shorter sentences	
positive emotion words	
social words	
self-references	
cheerful language	

Table 2. Agreeableness vs Disagreeableness Language Cues

After completing tasks 2 and 3 for the first chatbot, the participants repeat these steps for the second chatbot. To mitigate the order effect half of the participants are randomly selected to interact with Chatbot A first, while the other half will interact with Chatbot B and then Chatbot A. The questions in both post-interaction questionnaires are the exact same.

Task 4: Comparison Questionnaire. Lastly, participants complete a final post-interaction questionnaire. The objective of this questionnaire is to understand participants’ chatbot preference and to identify the differences they perceived between the chatbots. Participants were asked to describe any differences they noticed in their interactions with the chatbots in an open-text field. This was followed by a multiple choice question where participants chose their preferred chatbot. Finally, the participants explained this choice in an open-text field.

The questionnaires were carried out using Google Forms and the chatbots were connected to a web channel that would allow them to be accessed via link and used remotely while an online database container was used to store the experiment data.

A pilot study with four participants from the same population as the main study was conducted to evaluate comprehensibility of the questionnaire design, the conversation flow, and the experiment design. After this pilot, minor changes were made to the LUIS model including the addition of colloquial variations of

‘yes’ and ‘no’. Additionally, some questions posed by the chatbots were ambiguous due to the Covid-19 pandemic so these questions were edited and the virus was added as a topic of conversation. Participants reported confusion around the meaning of some words used in the pre-interaction personality questionnaire, such as ‘aloof’ and ‘reserved’. To prevent this confusion a list of definitions was supplied within the questionnaire for reference. The pilot study validated that the data collection was sufficient for addressing the research questions and the participants were excluded from the main study.

5 Results

Participant BFI personality scores for extraversion and agreeableness were calculated with a mean extraversion score of 3.06 ($\sigma = 0.63$) and a mean agreeableness score of 3.97 ($\sigma = 0.43$). We defined thresholds and grouped these scores to label participant personality traits: 5 participants were high in extraversion ($\text{score} \geq 3.5$), 10 participants were moderately extroverted ($2.5 < \text{score} < 3.5$), and two were low in extraversion ($\text{score} < 2.5$). Overall, the participants are high in agreeableness with 11 participants scoring ≥ 3.5 (high) and the remaining 6 participants scoring between 2.5 and 3.5 (moderate). All participants stated they knew what a chatbot was, 6 of whom described their understanding with technical detail. Their frequency of use was varied with 6 participants having never used a virtual assistant, 7 using them somewhat frequently, and 4 using them daily. So while all participants have a clear, and in some cases technical, mental model of chatbots, their experience using such agents is mixed.

To understand how participant behaviour varied between the agents, we analysed stylometric features in the interaction conversation logs. We calculated (i) the *duration* of the conversation in minutes using the timestamp from the first participant utterance until the timestamp of their last message, (ii) the number of participant *conversation turns*, (iii) the participant’s *total word count* which is the sum of the count of words in each utterance the participant submits to the bot, and (iv) the participant’s *mean utterance length* in words. See table 3 for descriptive statistics of these features for each agent. Overall, participants conversed more with Chatbot B than Chatbot A, across these engagement metrics. We ran a paired sample t-test for each measure and found the difference in behaviour captured by conversation duration ($p < 0.02$) and turn count ($p < 0.05$) were significant. Thus we can reject the null hypothesis that participants behaviour was the same across both agent interactions.

Participants were asked to rate the interaction with each agent on a Likert scale across 3 measures: knowledge, quality of conversation, and attitude/personality. Our null hypothesis states there is no difference in how participants perceive the personality of the two agents. Our alternative hypothesis states the participant discerns a difference in personality between the agents. We ran a paired sample t-test on each measure and found no statistically significant difference in user perception of agent knowledge ($p = 0.886$) or conversation quality ($p = 0.575$). However, we found a statistically significant difference in how users viewed per-

Measure	Chatbot	Minimum	Maximum	Mean	Variance
Duration (Minutes)	A	1	6	3.18	3.28
	B	2	21	6.35	18.37
Number of Turns	A	9	24	15.59	28.36
	B	10	34	20.29	35.38
Total Word Count	A	14	149	65.47	1951.66
	B	24	281	91.65	3263.64
Mean Utterance Length (Words)	A	7	36	18	86.67
	B	7	38	19	51.12

Table 3. Descriptive Statistics for Stylometric Features ($N=17$)

sonality ($p<0.02$). Given this evidence, we can reject the null hypothesis and determine that participants perceive a difference in personality of the agents. Participants were asked which agent they preferred interacting with; 12 chose Chatbot A and 5 chose Chatbot B. Interestingly, we did not find a statistically significant strong correlation between agent preference and participant personality trait scores. It is likely our sample is too small to capture any matching phenomenon that may exist.

We analysed the language used by participants in their descriptions of each chatbot for syntactic stylometric features (e.g. emoticons or expressive punctuation and language) to understand how they perceived the chatbots. NLTK¹¹, the NLP library for Python, was used to extract adjectives from the descriptions. The data was also manually analysed for adjectives not picked up by NLTK. Results are shown in Figure 3. Participants whose chatbot preference was Chatbot B described their experience with it as ‘engaging’, ‘personalised’, and ‘natural’. However, those who preferred Chatbot A described Chatbot B as ‘formal’, ‘robotic’, and the conversation as ‘unnatural’. These participants also described the experience as similar to *“being interviewed for RTE News”* (the national news service) or *“taking an oral exam when studying a language module”*. Similarly, those who preferred Chatbot B perceived Chatbot A as being ‘bland’, and ‘automated’ and felt the conversation was ‘not personalised’. Whereas participants who preferred Chatbot A cited its ‘cheery’, ‘bright’, ‘fun’, and ‘relaxed’ personality, with one participant saying Chatbot A had a *“better personality”* and another comparing the interaction to *“chatting with a friend”*. One participant perceived Chatbot A to be ‘nice’ but felt overall it was ‘bland’ and perceived Chatbot B as being more ‘engaged’ and thus preferred the interaction with Chatbot B. That same participant scored lower in both extraversion (2.57) and agreeableness (3) than the average participant, scores which match the personality traits of Chatbot B.

Lastly, the participants open-text responses across all post-interaction questionnaires were analysed for gendered language. A single participant used gender pronouns, specifically ‘he’ and ‘his’, but used these pronouns to describe both Chatbot A and Chatbot B. Interestingly, these pronouns were contrary to their own gender. Participants usually referred to the chatbots as ‘it’, by name, or as ‘chatbot’ or ‘bot’. This contrasts previous research that found users may still

¹¹ Natural Language Toolkit: <https://www.nltk.org/>

gender agents [24] without explicit visual cues. Our results may be due to limited relationship building with the agents or may be due to the participants well-developed mental model of chatbots.

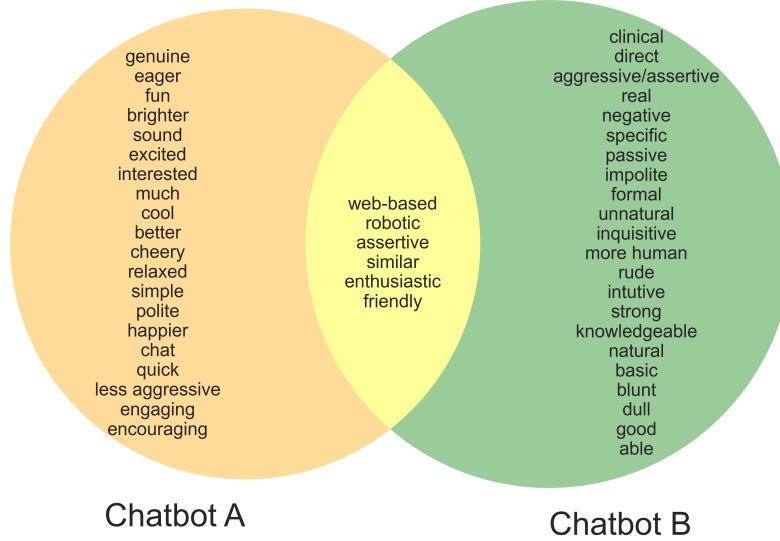


Fig. 3. Venn Diagram of adjectives used to describe Chatbot A and Chatbot B

6 Discussion

This section provides discussion of participant behaviour with, and attitudes towards, the chatbots based on the results presented in Section 5. We organize this discussion according to the two research questions outlined in Section 1.

6.1 RQ1: Can Personality Be Reliably Simulated by a Chatbot via Text Such That the User Perceives Personality or Personality Traits as Intended?

Our results support previous work (see Section 2) that suggest personality can be reliably simulated by a chatbot such that personality traits are perceived by users as intended. Participant descriptions of both chatbots are in line with the personality design; chatbot A was described as ‘friendly’ and ‘cheery’, reflecting high agreeableness and extraversion, while Chatbot B was described as ‘formal’ and ‘passive’, consistent with low agreeableness and low extraversion. Additionally, the agents were designed to vary on the axis of personality but not on their knowledge base or on the quality of conversation. When rating the chatbots along these three axes participants only perceived a difference in personality.

6.2 RQ2: Does the Perceived Personality of a Text-Based Chatbot Affect User Experience?

The results suggest the personalities of both chatbots had an effect on user experience. Participants showed higher engagement with Chatbot B than Chatbot A in terms of lexical and turn-taking features of conversation. We had hypothesised users would engage in longer conversations with their preferred chatbot but while overall participants spent more time conversing with Chatbot B the majority (70.6%) preferred Chatbot A. One reason for this behaviour may be a reaction to the direct nature and formality of language used by Chatbot B, something which many participants noted in their descriptions of it; “*Nasoto was pretty formal*”, “*very formal, not as friendly as makoto*”. In this case, participants may be mirroring the formality of the language used by the agent. In contrast, participants may have used shorter and more colloquial language to match the language style of Chatbot A. Such difference in behaviour of participants when interacting with the chatbots suggests the personality expressed through text of each chatbot had an overall effect on user experience. However, the effect itself is surprising and suggests (i) user engagement metrics may not be a good indicator of user preference for agent personality, and (ii) the application domain and goals of the agent should be considered when designing agent personality as the users’ generally preferred personality may not be the personality that leads to a productive user experience.

7 Limitations

To design two distinct personalities, we simultaneously varied extraversion and agreeableness. This is sufficient to answer our research questions. However, a 2x2 factorial design would allow for the measurement of the effect of each personality trait on user perception. Although our sample size of 17 is consistent with previous work in the literature, a larger sample size would increase the robustness and generalizability of the findings. Additionally, a larger sample size may contain more variance in personality traits among the user group, providing an opportunity to investigate how user personality affects user preference of agent personality in the context of a text-based conversational task (not strategic, collaborative, or commercial task).

8 Conclusion and Future Work

We have found that personality traits can be reliably simulated through text and can be perceived as intended by users without additional audio or visual cues. We have added to the body of evidence that agent personality impacts user experience and thus is an important design consideration. There is a lot of scope for future work that builds on this work and the outlined literature, including the use of more robust models of personality (rather than two dimensions of a five dimension model). Additionally, it would be of interest to investigate the effect

of domain on the perception of, and preference for, agent personality. A more emotional subject matter such as those covered by therapy chatbots may see very different user behaviour and attitudes than a chatbot that serves an information need, for example. Lastly, we analysed user perception of agent personality along the two dimensions we manipulated. It would be interesting to observe whether user perceptions of other personality traits differ even when those personality traits have not been explicitly varied. For example, do users perceive an increase in openness to experience (imaginative, spontaneous) when an agent is designed with high extraversion (sociable, fun-loving)?

9 Acknowledgements

This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094. Thank you to Dr. Brendan Rooney and Dr. Nicola Fox Hamilton for their very helpful advice. Thank you to Thomas Laurent and Glynda Ruane for their useful feedback on an earlier version of this manuscript.

References

1. Allbeck, J., Badler, N.: Toward representing agent behaviors modified by personality and emotion. *Embodied Conversational Agents at AAMAS* **2**, 15–19 (2002)
2. Berry, D.S., Pennebaker, J.W., Mueller, J.S., Hiller, W.S.: Linguistic bases of social perception. *Personality and Social Psychology Bulletin* **23**(5), 526–537 (1997)
3. Dewaele, J.M., Furnham, A.: Extraversion: The unloved variable in applied linguistic research. *Language Learning* **49**(3), 509–544 (1999)
4. Faur, C., Clavel, C., Pesty, S., Martin, J.C.: Perseed: a self-based model of personality for virtual agents inspired by socio-cognitive theories. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. pp. 467–472. IEEE (2013)
5. Gill, A.J., Oberlander, J.: Taking care of the linguistic features of extraversion. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. vol. 24 (2002)
6. Hanna, N., Richards, D., et al.: Do birds of a feather work better together? the impact of virtual agent personality on a shared mental model with humans during collaboration. In: *COOS@ AAMAS*. pp. 28–37 (2015)
7. Holtgraves, T.: Text messaging, personality, and the social context. *Journal of research in personality* **45**(1), 92–99 (2011)
8. Isbister, K., Nass, C.: Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International journal of human-computer studies* **53**(2), 251–267 (2000)
9. John, O.P., Srivastava, S., et al.: The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* **2**(1999), 102–138 (1999)
10. Kang, S.H., Gratch, J., Wang, N., Watt, J.H.: Agreeable people like agreeable virtual humans. In: *International Workshop on Intelligent Virtual Agents*. pp. 253–261. Springer (2008)
11. Lee, K.M., Nass, C.: Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology* **7**(1), 31–45 (2005)

12. Liew, T.W., Tan, S.M.: Virtual agents with personality: Adaptation of learner-agent personality in a virtual learning environment. In: 2016 Eleventh International Conference on Digital Information Management. pp. 157–162. IEEE (2016)
13. Liu, K., Tolins, J., Tree, J.E.F., Neff, M., Walker, M.A.: Two techniques for assessing virtual agent personality. *IEEE Transactions on Affective Computing* **7**(1), 94–105 (2015)
14. Luger, E., Sellen, A.: 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. pp. 5286–5297. ACM (2016)
15. Ma, X., Yang, E., Fung, P.: Exploring perceived emotional intelligence of personality-driven virtual agents in handling user challenges. In: The World Wide Web Conference. pp. 1222–1233 (2019)
16. Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research* **30**, 457–500 (2007)
17. McRorie, M., Sneddon, I., de Sevin, E., Bevacqua, E., Pelachaud, C.: A model of personality and emotional traits. In: International Workshop on Intelligent Virtual Agents. pp. 27–33. Springer (2009)
18. Moon, Y., Nass, C.I.: Adaptive agents and personality change: complementarity versus similarity as forms of adaptation. In: Conference companion on Human factors in computing systems. pp. 287–288 (1996)
19. Nass, C., Moon, Y., Fogg, B.J., Reeves, B., Dryer, C.: Can computer personalities be human personalities? In: Conference companion on Human factors in computing systems. pp. 228–229. ACM (1995)
20. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: International Conference on Intelligent Virtual Agents. pp. 222–235. Springer (2010)
21. Pennebaker, J.W., King, L.A.: Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology* **77**(6), 1296 (1999)
22. von der Pütten, A.M., Krämer, N.C., Gratch, J.: How our personality shapes our interactions with virtual characters - implications for research and development. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) *Intelligent Virtual Agents*. pp. 208–221. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
23. Roffo, G., Giorgetta, C., Ferrario, R., Cristani, M.: Just the way you chat: Linking personality, style and recognizability in chats. In: International Workshop on Human Behavior Understanding. pp. 30–41. Springer (2014)
24. Silvervarg, A., Haake, M., Gulz, A.: Educational potentials in visually androgynous pedagogical agents. In: International Conference on Artificial Intelligence in Education. pp. 599–602. Springer (2013)
25. Silvervarg, A., Raukola, K., Haake, M., Gulz, A.: The effect of visual gender on abuse in conversation with ecas. In: International Conference on Intelligent Virtual Agents. pp. 153–160. Springer (2012)
26. Smestad, T.L.: Personality Matters! Improving The User Experience of Chatbot Interfaces-Personality provides a stable pattern to guide the design and behaviour of conversational agents. Master's thesis, NTNU (2018)
27. Verhagen, T., Van Nes, J., Feldberg, F., Van Dolen, W.: Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication* **19**(3), 529–545 (2014)