On the Benefits of Directness in Virtual Characters for Motivational Interviews

Michael O'Mahony[1,2,*], Cathy Ennis[1,*] and Robert Ross[1,*]

[1]School of Computer Science, Technological University Dublin, Ireland
[2]SFI Centre for Research Training in Machine Learning at Technological University Dublin

Abstract

Understanding the factors influencing successful engagement with Embodied Conversational Agents (ECAs) remains a significant challenge. This understanding could be used to personalise agents to users to improve interactions. Some studies have shown that simulating personalities in healthcare agents can improve effectiveness and engagement. However, it is not yet well understood how variations of agent personality can be leveraged to improve user engagement with Motivational Interviewing (MI) ECAs. Specifically how the balance between agent warmth and directness can be controlled in an MI agent to improve likeability and engagement. We conducted an online Wizard-of-Oz (WoZ) mediated study of two variants of a motivational ECA to investigate user perception and attitudes towards warmth and directness in interaction style. In our MI scenario, participants rated likeability and engagement higher for the direct agent variation. This effect was not as strong for younger participants or participants who were not native English speakers. This result gives us a direction to improve MI ECAs to make their increased adoption more likely.

Keywords

Embodied Conversational Agents, Text Style Control, Motivational Interviewing, Personality, Text Generation

1. Introduction

With recent advancements in Large Language Models (LLMs), conversational systems have been applied to many more tasks across multiple domains, including customer support, language learning, and healthcare. Healthcare agents can alleviate pressure on overburdened healthcare systems and provide users with access to care who may not have access otherwise due to financial or geographical reasons. Personalising aspects of these agents to users can improve user satisfaction and engagement [1, 2, 3], which leads to interactions with these agents lasting longer or occurring more frequently.

Many works have investigated the personalisation of healthcare Conversational Agents (CAs). Usually, the personalised aspect is the content of the generated text [3]. Another aspect which could be personalised is the style of the generated text. The style of generated text is how something is said rather than what is said, as there are often many ways to say the same thing. It is an essential aspect of conversational systems as many applications require that information is delivered a certain way and it plays a significant role in the user satisfaction of a dialogue system [4]. Recent advances in LLMs have made Text Style Control (TSC) much more accessible and generalisable through prompt engineering and In-Context Learning (ICL). Some studies have shown that it is possible to imitate personality through text style [5, 6, 7], which opens up many possibilities to improve aspects of human-agent interactions, including engagement and likeability, without changing the content delivered.

Motivational Interviewing (MI) is a counselling technique used to increase the motivation of a participant to change their behaviour. It is one of the most effective psychological interventions for this purpose [8]. Multiple studies have investigated the ability of virtual agents to deliver MIs to participants, and validated their effectiveness [9, 10, 11, 12]. Yet there is a lack of work which examines the issues of personality and style in the context of MI agents.

AICS'24: 32nd Irish Conference on Artificial Intelligence and Cognitive Science, December 09–10, 2024, Dublin, Ireland
*Corresponding author.
michael.t.omahony@mytudublin.ie (M. O'Mahony); cathy.ennis@tudublin.ie (C. Ennis); robert.ross@tudublin.ie (R. Ross)
0000-0003-2344-7377 (M. O'Mahony); 0000-0002-1274-5347 (C. Ennis); 0000-0001-7088-273X (R. Ross)

We designed an online user study of two variants of a motivational Embodied Conversational Agent (ECA) to investigate user perception of and attitudes towards warmth and directness in interaction style. As some works have shown that varying personality in MI ECAs can improve the effectiveness of the MI intervention [11]. Since this was a Wizard-of-Oz (WoZ) mediated study, the agent's dialogue turns were initiated by the researcher. We used ChatGPT to change the text style in the agent's script to simulate warm and direct personalities. Someone who has a warm personality is considered friendly and invested in others [13]. Directness refers to the degree to which information is communicated concretely [14]. We recruited participants from local communities and evaluated agent likeability and user engagement using objective metrics and a general agent rating questionnaire [10, 11]. We also collected user personality data using the Ten-Item Personality Inventory (TIPI) [15], and other participant information. This work reports the details of our experiment, data collection, and analysis of how agent personality affected the agent rating metrics, and the extent to which user personality, among other factors, could be used to predict preferred agent personality to personalise agents to users.

In summary, our contributions are as follows:

1. A WoZ human-avatar interaction study design for investigating warm and direct personality variations in an ECA.
2. The analysis result of the differences in questionnaire responses between participant groups to outline a direction for improving MI ECAs.
3. The pseudonymised dataset containing participant demographic and personality information, evaluation questionnaire responses, and interaction audio, which we will make public to promote further research into personality variants of MI ECAs[1].

2. Related Work

A Motivational Interview (MI) is a counselling technique used to motivate participants to change unwanted behaviour. Multiple studies have employed virtual agents or robots to deliver MIs to participants to increase their motivation to eat healthier [10, 11], exercise more [10, 11, 8, 9, 16], or reduce alcohol consumption [12]. A few studies focus on agent personality variations, often empathy [2, 16], but there are also works focused on other aspects, including humour [11] and adaptivity [1]. Although there are many nuanced factors when designing ECAs with simulated personalities, these studies have demonstrated that the personality of a virtual agent can affect its likeability, engagement, or effectiveness. Olafsson et al. (2020) [11] found that participants who interacted with their humorous ECA had a significantly greater change in motivation than when they interacted with the non-humorous agent, and when they were given the option to have another conversation, most (12 versus 3) selected the humorous agent. Barange et al. (2022) [2] reported that their adaptive empathic ECA engaged users more. Chauvin et al. (2023)'s [16] results were more mixed. Many of the authors' results did not reach the threshold of statistical significance. However, Chauvin et al.'s empathic ECA was perceived as more credible and empathic, and it increased some types of participant self-efficacy and motivation than their MI ECA without empathy, and text-based website. While simulated agent personality has the potential to improve interactions with users, what may be an improvement for some may be a dis-improvement for others. Personalised agents have the potential to deliver the right aspects to the right people and improve the interactions overall. To achieve this goal, more work is needed to understand the subtleties of personality variation in MI ECAs and their relationship to engagement.

The personalisation of CAs in healthcare has been increasing in use [3]. Personalisation can be used to improve numerous aspects of CAs such as user comprehension, user satisfaction, task-efficiency, and the likelihood of behaviour change. There are many ways healthcare CAs can be personalised, but it is usually the content [3]. Barange et al. (2022)'s [2] adaptive high-level empathic ECA was more engaging and was perceived as more empathic than their non-adaptive low-level empathic ECA. Egede et al. (2021) [1] conducted a WoZ-mediated user study to research whether an adaptive ECA was more

[1]Data will be available at https://github.com/MichaelOMahony/getting-to-the-point-data/ after 01/01/2025.

Table 1
General Agent Ratings [10, 11]

Number	Question
1	I am satisfied with the agent
2	I would continue talking with the agent
3	I trust the agent
4	I like the agent
5	The agent was knowledgeable
6	The conversation was natural
7	I have a good relationship with the agent
8	I am similar to the agent

engaging than a non-adaptive ECA or an adaptive non-embodied CA when delivering health advice to pre and post-natal maternal women. The results revealed participants were more engaged with the adaptive ECA. However, many nuanced factors affected engagement, and further work was called for to understand these nuances.

One way in which personalization can be achieved is through Text Style Control (TSC) which is a sub-field of Text Generation. It is the practice of maintaining a desired style consistently when generating text, where the style of text how the information is conveyed rather than what the information is. For example, “I did an experiment" is a more casual way to say, “The experiment was carried out by me", while the two sentences have the same meaning. The Text Generation field has increased in interest over time [17]. Advancements in neural networks, the transformer architecture, and LLMs have prompted significant growth in the field. Although most of this work has been on the content fidelity of generated text rather than style, text style has seen increased interest in recent years [17, 18].

The main strategies for controllable text generation using LLMs involved re-training, fine-tuning, or post-processing. These methods are costly financially and time-wise, either at the training or inference stages, especially with the increasing number of LLM parameters [19]. These methods also require a level of expertise. However, more recent LLMs of sufficient size are believed to have learned a large amount of semantic and syntactic knowledge from massive amounts of data and can generate text of unprecedented quality [19]. It also has become increasingly possible to control the style of generated text through In-Context Learning (ICL) [20]. ICL involves providing training samples to the model at inference time using the prompt. ICL does not require any re-training or fine-tuning, does not require access to model weights, and has been shown to significantly improve performance on downstream tasks [21]. For dialogue systems, it is often required to control aspects such as emotion, persona, or politeness through the style of the generated text [19].

Various studies have shown that it is possible to control the perceived personality of generated text based on its style [6, 7, 5], but it is not yet known what the impact is on user engagement or agent likeability in the context of MI ECAs. Many of the state-of-the-art works investigating personality variations of ECAs in healthcare designed these variations by differing the content delivered during interactions. Our work is different as we aim to influence the likeability and engagement of an ECA by changing the text style only without adding new content. Some related works focus on empathy [2, 16]. However, as empathy is challenging to simulate through text style control only, our work focuses on agent warmth. Warmth is an aspect of personality which is known to be desired in human counsellors [22], but the extent of which is not fully agreed upon in the field of psychology [23]. There has not been extensive work done on warmth in ECAs, but it has been demonstrated that it has a positive effect on agent believability [24] which can enhance interactions with virtual agents. We selected directness as an alternative personality to contrast warmth.

3. Methodology

In our previous work [25], we introduced the outline of the methodology and initial results of an online WoZ user study. We designed the study to research the perception and impact of warmth and directness variation in MI ECAs. Initial results suggested that agent directness may be preferred; however, there was insufficient evidence to fully verify this preference. In this paper, we will give a more detailed overview of the study and provide a detailed analysis.

Our hypothesis is the following:

> From the results of the general agent ratings (see table 1), there will be a preference for the “warm" over the “direct" agent personality.

As perceived personality can be controlled through the style of text [6, 7, 5], and the style of text is an important factor in the user satisfaction of a dialogue system [4].

3.1. Experiment Design

The interaction scenario was an MI delivered by a virtual agent to increase users’ motivation to change their exercise behaviour. 25 participants were recruited. They were asked to use their own computer, in a quiet space to themselves, with a stable internet connection. Participants could listen using headphones or speakers and would indicate their choice in the pre-interaction questionnaire. Participants would interface with the ECA via voice though mediated through an online interface. The MI script was adapted from an earlier study [8]. Galvão Gomes Da Silva et al. (2018) [8] designed the script so that each question should make sense to the user, irrespective of how they answered previous questions. Therefore, multiple dialogue branches to handle each potential user response did not need to be designed. Building on the existing corpus, we created two conditions:

- A: Warmer personality
- B: More direct personality

by altering parts of the original script using ChatGPT. In practice, we only changed the beginning and end of the original script, aside from a minor manual change in the first question for clarity, we did not alter any of the questions as designing an MI was outside the scope of the work. This meant that only a small fraction of the interaction was changed across the two groups, but this is a first look at the impact of these personality variations. The agent asked questions to help increase participants’ motivation to positively change their exercise habits. In general, each experiment lasted 20-30 minutes, with the MI interaction lasting between 5 and 15 minutes.

The agent was given a virtual appearance using the Unity game engine[2], along with a Ready Player Me avatar[3] as they are industry standard, and widely used. Moreover, we used the Talking With Hands live motion capture dataset [26] for the talking gestures, Ready Player Me animation library[4] for the idle animation, and Salsa Lip Sync[5] to generate realistic mouth animations when the agent was talking. Salsa Lip Sync was used because we deployed our system to the WebGL library so users could participate online via a web browser, without having to download and install an executable file. Google’s Cloud AI[6] text-to-speech was used for the agent’s voice, where we selected a female avatar, regardless of the participant’s gender as some studies suggest that men slightly prefer a female therapist to a male one or do not care, and women are much more likely to prefer a female therapist [27, 28]. As the MI ECA followed a set script, the talking animations and speech were pre-set. Every participant was asked the same questions and saw the same animations.

[2] https://unity.com/
[3] https://readyplayer.me/
[4] https://github.com/readyplayerme/animation-library
[5] https://crazyminnowstudio.com/unity-3d/lip-sync-SALSA/
[6] https://cloud.google.com/products/ai/

Figure 1: MI ECA Appearance

3.2. Dialogue Design

As indicated earlier, the virtual agent followed a set script and was controlled through a WoZ. There were no options to change the next utterance based on the participants' responses, but we could repeat the last question if it was requested. The script was adapted from a work that also delivered MIs to participants [8], but they used a NAO robot rather than an ECA. In the original study, the participant controlled when the next utterance was delivered by pressing a button on the robot's head. We used a WoZ setup to simulate a more natural conversation. The MI was delivered to participants to increase their motivation to change their exercise behaviour in the way they believe to be a positive direction. It did not prescribe actions participants should take to improve their behaviour, but asked questions that led them to say what they believe they should change out loud in their own words. MI is one of the most effective psychological techniques for helping to change the behaviour of a participant, including exercise behaviour [8].

Galvão Gomes Da Silva et al. [8] designed the script so each question should make sense, independently of how the participant answered the last. This way, multiple possible responses did not need to be designed, and the participant could answer quite openly, in contrast to a lot of the similar works in this area that constrain the user to selecting one out of a few options [11, 10, 16]. In practice, this method mostly worked, but there were instances where a somewhat broad question led to some confusion on the participant's part.

The prompts we used with ChatGPT to generate the “warmer" and “more direct" versions were:

1. “In the following I will give you a number of utterances. Please rewrite these points to keep the original intent but make the language warmer and more friendly".

2. “In the following I will give you a number of utterances. Please rewrite these points to keep the original intent but make the language more direct and less warm".

Minor modifications were made to the original script based on test runs to improve clarity.

3.3. Questionnaires and Recruitment

Participants answered questionnaires before and after interactions, and interaction timestamps and audio recordings were collected. The pre-interaction questionnaire included demographics, exercise frequency, the TIPI for the participant [15], familiarity and acceptance of virtual agents. The post-interaction questionnaire included the TIPI for the virtual agent [15], general agent ratings [10, 11] (see table 1) and an open-ended feedback box. The TIPI considers five personality dimensions, “The Big Five", which are openness, conscientiousness, extraversion, agreeableness, and neuroticism. This is the most widely accepted model of personality in practice. We lightly altered the wording of the general agent ratings to use the same response scale for each one, which participants answered using a five-point Likert scale. Each personality dimension was calculated by averaging two 5-point Likert scale values, and each question response is a 5-point Likert scale value.

To measure whether participants were considered active or not for further analysis, we asked them their frequency of exercising at mild, moderate, and vigorous intensity in the pre-interaction questionnaire. The Irish Health and Safety Executive (HSE) guidelines state that adults 18-64 years old should exercise at moderate intensity for at least 30 minutes a day for five days a week. The United States Department of Health recommends at least 150 minutes of moderate-intensity activity per week, 75 minutes of vigorous intensity per week, or a mixture of both. We defined a participant as “active" if they exercised at moderate intensity “4-6 times per week" or more, or vigorous intensity “2-3 times per week" or more. There were 9 participants considered relatively active and 16 participants considered relatively inactive based on these guidelines.

We recruited 25 participants in total using email mailing lists to staff and postgraduate students in our School of Computer Science, as well as outside social networks. Of these, 12 were male, 12 were female, and 1 was non-binary. We recruited 4 participants in the 18-24 age range, 13 participants in the 25-34 age range, 0 in the 35-44 age range, 4 in the 45-54 age range, and 4 in the 55 and over age range. For further analysis, we grouped these into two groups, “younger": 18-34 years old, and “older" 45 years and over. 16 participants were native English speakers, and 9 were not. There were 10 unique native languages among all the participants, but all were fluent in English. Two of the native English speakers were also native speakers of one other language. Our inclusion criteria was participants had to be over 18 years old, and speak English fluently.

To control for participants' prior attitudes towards virtual agents, we included three questions in our pre-interaction questionnaire, which were adapted from questions proposed by a modified Technology Acceptance Model [29]:

1. I think it is a good idea to use this technology to increase my motivation.
2. I think that this technology will be easy to use.
3. I would use this technology if it became available to me.

4. Results

In our prior work, we reported the results of each question analysed individually (see table 2, figure 2). Only Q7 was statistically significantly different in favour of the direct version (2.85 versus 3.42, p=0.0379, U=42.0). We computed these results using Mann-Whitney U hypothesis tests, as most samples were not normally distributed, according to Shapiro-Wilk tests. However, it is essential to consider all of the general agent ratings together to paint a complete picture of the effects of the warmth and directness personality variation in our interaction scenario. We use two-way ANOVA tests to analyse the effects of agent personality and question number on question responses. We also analyse the effects of demographics, exercise frequency, personality, familiarity, and acceptance towards virtual agents on the results. An understanding of these relationships could be used to personalise virtual agents to users.

As reported in our prior work [25], the mean responses for all eight questions (see table 1) were higher from participants in the direct group than participants in the warm group (see figure 2). When analysed on a question-by-question basis, the only question which demonstrated a statistically significant difference was Q7 “I have a good relationship with the agent". The results of the two-way ANOVA, however (see table 3), reveal a significant (p=0.0336, F=4.56) positive effect of the direct personality variation compared to the warm variation on the results of the general agent ratings.

Considering how each category impacted the individual ratings, male participants rated the direct agent higher than the warm agent (means: direct = 4.50, warm = 3.63) for Q4: “I like the agent", and this result approached significance (p=0.0543, U=6.0). Female participants did not have a significant

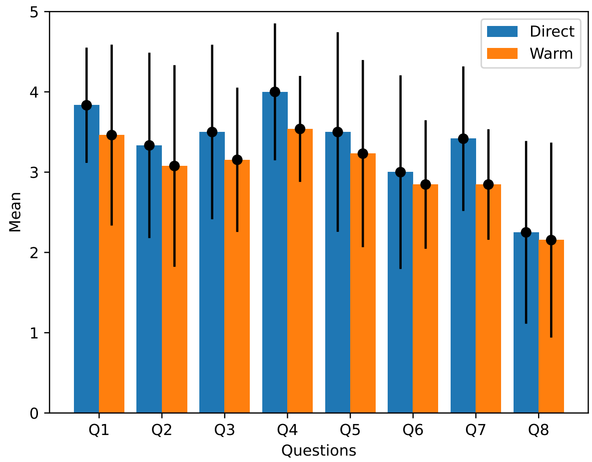

Figure 2: Mean Responses to the General Agent Ratings [10, 11], with Standard Deviation Bars

Table 2
Means and p-values (Mann Whitney U test) for Each Question

Number	Question	Warm	Direct	Difference	p	U
1	I am satisfied with the agent	3.46	3.83	-0.37	0.4494	66.0
2	I would continue talking with the agent	3.08	3.33	-0.26	0.5729	67.5
3	I trust the agent	3.15	3.50	-0.35	0.4058	63.0
4	I like the agent	3.53	4.00	-0.46	0.0953	50.5
5	The agent was knowledgeable	3.23	3.50	-0.27	0.5685	67.5
6	The conversation was natural	2.85	3.00	-0.15	0.5495	67.0
7	**I have a good relationship with the agent**	**2.85**	**3.42**	**-0.57**	**0.0379****	**42.0**
8	I am similar to the agent	2.15	2.25	-0.10	0.8193	73.5

Table 3
Two-Way ANOVA Results

Variable	Sum of Squares	df	F	$PR(> F)$
is_warm	4.98	1.00	4.58	**0.0336****
question	21.78	1.00	20.04	0.00001*
Residual	214.03	197.00	NaN	NaN

preference between the warm and direct agents for Q4 (means: direct = 3.75, warm = 3.5, p=0.5037, U=12.0). According to a three-way ANOVA, considering agent warmth/directness, question number, and gender; the interaction between gender and question number is a significant factor in predicting question responses (p=0.0340, F=3.4418).

In a two-way ANOVA with age group and question number as independent variables, age group approaches significance (p=0.0577, F=3.6442). The interaction between age group and question number is also statistically significant (p=0.0359, F=4.4629). A three-way ANOVA, which also considered the agent warmth/directness as another variable, reveals warmth/directness (p=0.0462, F=4.0257) and the interaction between age group and question number (p=0.0334, F=4.5881) to be significant. Therefore, different age groups experienced the interaction differently. Younger participants (18-34) seemingly had an inconclusive opinion on the two agent variations. This is verified by a two-way ANOVA test, considering whether the variation was warm or direct and the number of questions with responses from the younger group. The p-value for the warm/direct variable was 0.3608 (F=0.8397), showing that it did not have a significant effect on the responses to the ratings from the younger participants. However, older participants (45+) in the direct group had significantly higher responses than the warm

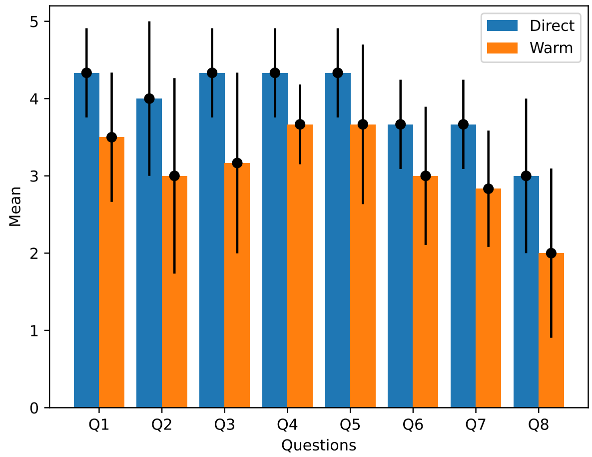

Figure 3: Mean Responses to the General Agent Ratings [10, 11], with Standard Deviation Bars (Relatively Active Group)

Table 4
Personality Analysis Regression Select Results

Question	Agent Personality	Personality Dimension	Coefficient	$P > \|t\|$	t
Q1	Direct	Extraversion	-0.55	0.078	-2.127
Q6	**Direct**	**Agreeableness**	**1.55**	**0.035***	**2.705**
Q6	Direct	Extraversion	-0.77	0.099	-1.949
Q7	Direct	Agreeableness	1.08	0.079	2.113
Q7	Direct	Openness	1.51	0.101	1.934
Q8	Warm	Extraversion	0.75	0.088	1.978
Q8	Direct	Openness	2.06	0.106	1.903

group for Q4 (means: 4.25 versus 2.75, p=0.0228, U=16.0) and Q7 “I have a good relationship with the agent" (3.75 versus 2.25, p=0.0325, U=15.5).

Agent warmth/directness (p=0.0331, F=4.6049), the interaction between agent warmth/directness and whether the participant was a native English speaker (p=0.0106, F=6.6540), and question number (p=0.0017, F=10.1697) are all significant variables in a three-way ANOVA with question responses as the dependent variable, revealing that native and non-native speakers experienced the personality variants of the MI agent differently. Non-native English speakers were more likely to prefer the warm agent than native English speakers. A two-way ANOVA test verifies this claim, considering all the questions together with a p-value of 0.0387 (F=4.3870) for the binary native English variable, and that the mean for every question was higher. However, individually, only Q7 was significant (means: 3.50 versus 2.56, p=0.0327, U=5.0). Though Non-native English speakers had a mixed opinion of the warm agent compared to the direct agent (two-way ANOVA warm/direct variable p=0.4312, F=0.6242). Responses from participants who were considered relatively active had higher means for all questions except for Q7 (which was close (active = 3.111, inactive = 3.125)) compared to relatively inactive participants (see figure 3). When measured individually, none of the differences were statistically significant. However, when tested together, the two-way ANOVA results reveal a significant effect of whether the participant was considered relatively active on the question responses (p=0.0446, F=4.0860). The relatively active group’s responses were higher for the direct condition for every question when compared to the warm condition, but these were not significant individually, meaning relatively active participants may have had a better overall experience of both ECA variations.

To investigate the effects of participant personality and agent personality on preferred agent personality, we used multiple linear regression models to consider the effects of all five dimensions of participant personality together with agent personality (warmth and directness) on question responses and to investigate the importance of individual features when all features are considered together. These results were largely statistically insignificant. To understand how participant personality dimensions influenced preferences of agent personalities, we measured correlations between individual personality dimensions and question responses. We also used linear regression models for each question, considering either warm or direct observations at a time. The analysis of the regression model outputs reveals one minor significant result, and more results which were approaching significance (see table 4). For Q6: "The conversation was natural", agreeableness is a significant positive predictor for the direct variation (coefficient (c)=1.55, p=0.035). Agreeableness is also a (non-significant) positive predictor for the direct agent for Q7: "I have a good relationship with the agent" (c=1.08, p=0.079). Extraversion is a (non-significant) negative predictor for the direct ECA for Q1: "I am satisfied with the agent" (c=-0.5525, p=0.0780), for Q6 (c=-0.7730, p=0.0990), and a (non-significant) positive predictor for the warm agent for Q8: "I am similar to the agent" (c=0.7524, p=0.0880). Openness is a (non-significant) positive predictor for the direct variation for Q7 (c=1.5127, p=0.101) and Q8 (c=2.0646, p=0.1060). In general, there appears to be a trend towards extraverted participants preferring the warm over the direct agent, but a higher sample size and more research into the subtle relationship between participant personality and preferred ECA personality is needed to verify this possible effect. To evaluate the effects of participant acceptance towards virtual agents on their responses, we used multiple linear regression models considering responses to the three acceptance questions and whether the agent variation was warm or direct. The results of the analysis of the acceptance measures are largely statistically insignificant.

5. Discussion

The results demonstrate that the direct variation of our ECA had a positive influence on interactions with participants. Therefore we reject our hypothesis: 'From the results of the general agent ratings (see table 1), there will be a preference for the "warm" over the "direct" agent personality', as the two-way ANOVA test results reveal that whether the agent was the warm or direct variation had a statistically significant effect on the question responses, and the mean responses for every question are higher for the direct agent. We can conclude that there was a general preference for the direct agent variation for our scenario. The general preference for the direct variation may be because participants prefer a more direct virtual counsellor for the scenario of MI or because it was slightly shorter and got to the point faster. This effect may be related to an older study that found while empathy, warmth, and genuineness together were desired traits in human counsellors, warmth without the other two negatively affected outcomes [30]. It is unknown whether this observed effect in human counsellors applies to ECA counsellors. Participants may have found the warm ECA ingenuine. A follow-up study is needed to verify whether this was the case.

ANOVA tests reveal that younger participants (18-34) and non-native English speakers are not statistically significantly affected by the agent's personality variations. The observed effect may only be from older participants (45+) or native English speakers. Therefore, we propose that the observed preference for directness over warmth is considered when designing MI ECAs, especially for older participants and native English speakers. However, more work is needed to explore and understand this effect fully. Non-native English speakers may be more likely to prefer the warm variation than native English speakers because the language is easier to understand. Active participants may be more likely to prefer the interactions as they may be more comfortable conversing about their exercise habits than less active participants. However, more work is required to validate these assumptions.

Due to the small sample size, and to the sample being further divided when we analysed the effects of different categories on the results, some results did not reach statistical significance, which may have if we had more participants. As designing a counselling intervention was outside the scope of our experiment, we did not alter the questions when changing the text style of the script to simulate agent personalities. We only changed the start and the end of the script. If we had changed the whole script, the effect of personalities on the interaction may have been stronger. Most of our participants were in the 25-34 age range (13), and we did not recruit any participants in the 35-44 age range. While the results are promising, we did not conduct false discovery rate corrections. Therefore, the significance of some of the results may not be sufficient. Nonetheless, we have highlighted encouraging directions for future research.

6. Conclusions and Future Work

We designed and conducted an online WoZ-mediated user study to investigate participants' preferences for warm and direct personality variations of an MI ECA. We simulated these personalities by altering the text style of an MI script, adapted from another study [8] using ChatGPT. Participants answered a general agent rating questionnaire [10] (see table 1) after their interaction. Based on user ratings, we conclude that the direct agent variation is preferred. However, this preference may only be held by older (45+) or native English-speaking participants.

Future work will focus on nuancing the qualities of directness and warmth in speech and embodying these in a more automated agent with evaluation of effectiveness as well as engagement.

Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6183. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[1] J. Egede, M. J. G. Trigo, A. Hazzard, M. Porcheron, E. Bodiaj, J. E. Fischer, C. Greenhalgh, M. Valstar, Designing an adaptive embodied conversational agent for health literacy: a user study, in: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 112–119. URL: https://doi.org/10.1145/3472306.3478350. doi:10.1145/3472306.3478350.

[2] M. Barange, S. Rasendrasoa, M. Bouabdelli, J. Saunier, A. Pauchet, Impact of adaptive multimodal empathic behavior on the user interaction, in: Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents, IVA '22, Association for Computing Machinery, New York, NY, USA, 2022. URL: https://doi.org/10.1145/3514197.3549675. doi:10.1145/3514197.3549675.

[3] A. Kocaballi, S. Berkovsky, J. Quiroz, L. Laranjo, H. Tong, D. Rezazadegan, A. Briatore, E. Coiera, The personalization of conversational agents in health care: Systematic review, Journal of Medical Internet Research 2019;21(11):e15360 (2019). doi:10.2196/15360.

[4] B. Peng, C. Zhu, C. Li, X. Li, J. Li, M. Zeng, J. Gao, Few-shot natural language generation for task-oriented dialog, in: Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, p. 172–182.

[5] M. Li, H. Liu, B. Wu, T. Bai, Language style matters: Personality prediction from textual styles learning, in: 2022 IEEE International Conference on Knowledge Graph (ICKG), 2022, pp. 141–148. doi:10.1109/ICKG55886.2022.00025.

[6] S. Argamon, S. Dhawle, M. Koppel, J. Pennebaker, Lexical predictors of personality type, in: 2005 Joint Annual Meeting of the Interface and the Classification Society of North America, 2005. Place of conference:USA.

[7] F. Mairesse, M. Walker, PERSONAGE: Personality generation for dialogue, in: A. Zaenen, A. van den Bosch (Eds.), Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 496–503. URL: https://aclanthology.org/P07-1063.

[8] J. Galvão Gomes da Silva, D. J. Kavanagh, T. Belpaeme, L. Taylor, K. Beeson, J. Andrade, Experiences of a motivational interview delivered by a robot: Qualitative study, in: J Med Internet Res, 2018.

[9] J. Galvão Gomes da Silva, D. J. Kavanagh, J. May, J. Andrade, Say it aloud: Measuring change talk and user perceptions in an automated, technology-delivered adaptation of motivational interviewing delivered by video-counsellor, Internet Interventions 21 (2020) 100332. URL: https://www.sciencedirect.com/science/article/pii/S2214782920300981. doi:https://doi.org/10.1016/j.invent.2020.100332.

[10] S. Olafsson, T. O'Leary, T. Bickmore, Coerced change-talk with conversational agents promotes confidence in behavior change, in: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth'19, Association for Computing Machinery, New York, NY, USA, 2019, p. 31–40. URL: https://doi.org/10.1145/3329189.3329202. doi:10.1145/3329189.3329202.

[11] S. Olafsson, T. K. O'Leary, T. W. Bickmore, Motivating health behavior change with humorous virtual agents, in: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, IVA '20, Association for Computing Machinery, New York, NY, USA, 2020. URL: https://doi.org/10.1145/3383652.3423915. doi:10.1145/3383652.3423915.

[12] S. Olafsson, P. Pedrelli, B. C. Wallace, T. Bickmore, Accomodating user expressivity while maintaining safety for a virtual alcohol misuse counselor, in: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents, IVA '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3570945.3607361. doi:10.1145/3570945.3607361.

[13] P. Costa, R. McCrae, The revised neo personality inventory (neo-pi-r), The SAGE Handbook of Personality Theory and Assessment 2 (2008) 179–198. doi:10.4135/9781849200479.n9.

[14] J. Miehle, W. Minker, S. Ultes, What causes the differences in communication styles? a multicultural study on directness and elaborateness, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, 2018. URL: https://aclanthology.org/L18-1625.

[15] S. D. Gosling, P. J. Rentfrow, W. B. Swann, A very brief measure of the big-five personality domains, Journal of Research in Personality 37 (2003) 504–528. URL: https://www.sciencedirect.com/science/article/pii/S0092656603000461. doi:https://doi.org/10.1016/S0092-6566(03)00046-1.

[16] R. Chauvin, C. Clavel, N. Sabouret, B. Ravenet, A virtual coach with more or less empathy: impact on older adults' engagement to exercise, in: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents, IVA '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3570945.3607338. doi:10.1145/3570945.3607338.

[17] Y. Su, D. Vandyke, S. Wang, Y. Fang, N. Collier, Plan-then-generate: Controlled data-to-text generation via planning, in: Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, p. 895–909.

[18] S. Lin, W. Wang, Z. Yang, X. Liang, F. F. Xu, E. Xing, Z. Hu, Data-to-text generation with style imitation, in: Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, pp. 1589–1598.

[19] H. Zhang, H. Song, S. Li, M. Zhou, D. Song, A survey of controllable text generation using transformer-based pre-trained language models, ACM Comput. Surv. 56 (2023). URL: https://doi.org/10.1145/3617680. doi:10.1145/3617680.

[20] N. Wies, Y. Levine, A. Shashua, The learnability of in-context learning, in: Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23, Curran Associates Inc., Red Hook, NY, USA, 2024.

[21] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, volume 33, Curran Associates, Inc., 2020, pp. 1877–1901. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

[22] D. A. Shapiro, Empathy, warmth and genuineness in psychotherapy,

British Journal of Social and Clinical Psychology 8 (1969) 350–361. URL: https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8260.1969.tb00627.x. doi:https://doi.org/10.1111/j.2044-8260.1969.tb00627.x. arXiv:https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2044-8260.1969.tb

[23] C. H. Patterson, Empathy, warmth, and genuineness in psychotherapy: A review of reviews, Psychotherapy: Theory, Research, Practice, Training, 21(4) (1984) 431–438. doi:https://doi.org/10.1037/h0085985.

[24] R. Niewiadomski, V. Demeure, C. Pelachaud, Warmth, competence, believability and virtual agents, in: J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, A. Safonova (Eds.), Intelligent Virtual Agents, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 272–285.

[25] M. O'Mahony, C. Ennis, R. Ross, Getting to the point: Contrasting directness and warmth in motivational embodied conversational agents, in: Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts, SEMDIAL, Trento, Italy, 2024. URL: http://semdial.org/anthology/Z24-OMahony_semdial_0025.pdf.

[26] G. Lee, Z. Deng, S. Ma, T. Shiratori, S. Srinivasa, Y. Sheikh, Talking with hands 16.2m: A large-scale dataset of synchronized body-finger motion and audio for conversational motion analysis and synthesis, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 763–772. doi:10.1109/ICCV.2019.00085.

[27] L. Liddon, R. Kingerlee, J. A. Barry, Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking, British Journal of Clinical Psychology 57 (2018) 42–58. URL: https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1111/bjc.12147. doi:https://doi.org/10.1111/bjc.12147. arXiv:https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/bjc.12147.

[28] Z. E. Seidler, M. J. Wilson, D. Kealy, J. L. Oliffe, J. S. Ogrodniczuk, S. M. Rice, Men's preferences for therapist gender: Predictors and impact on satisfaction with therapy, Counselling Psychology Quarterly 35 (2022) 173–189. URL: https://doi.org/10.1080/09515070.2021.1940866. doi:10.1080/09515070.2021.1940866. arXiv:https://doi.org/10.1080/09515070.2021.1940866.

[29] M. P. Gagnon, E. Orruño, J. Asua, A. B. Abdeljelil, J. Emparanza, Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system, Telemedicine journal and e-health : the official journal of the American Telemedicine Association 18(1) (2012) 54–59. URL: https://doi.org/10.1089/tmj.2011.0066.

[30] C. B. Truax, D. G. Wargo, J. D. Frank, S. D. Imber, C. C. Battle, R. Hoehn-Saric, E. H. Nash, A. R. Stone, Therapist empathy, genuineness, and warmth and patient therapeutic outcome, Journal of Consulting Psychology, 30(5) (1966) 395–401. doi:https://doi.org/10.1037/h0023827.