Advantages of Virtual Agents over Clinical Psychologists during Comprehensive Mental Health Interviews Using a Mixed Methods Design

Kenji Yokotani

Niigata Seiryo University

Gen Takagi, Kobun Wakashima

Tohoku University

Author Note

Kenji Yokotani is with the Graduate School of Clinical Psychology, Niigata Seiryo University, Niigata city, Japan. Gen Takagi and Kobun Wakashima are with the Graduate School of Education, Tohoku University, Sendai city, Japan.

The present study was funded by a grant from the Foundation for the Fusion of Science and Technology.

Correspondence concerning this article should be addressed to Kenji Yokotani, Niigata Seiryo University, Graduate School of Clinical Psychology, 1-5939, Suidocho, Chuo-ku, Niigata-shi, Niigata 951-8121, Japan. Phone: 0081-25-266-0127; Fax: 0081-25-267-0053; E-mail: yokotani@n-seiryo.ac.jp

Abstract [1]

The use of Virtual Agents (VAs) is currently a popular topic in mental health interviews. Advantages of VA over Real Expert (RE) in the interview were reported. However, the advantages of audio-visual VAs over REs during comprehensive mental health interviews remain unclear, and their clarification is important to promote the practical application of VAs in these settings. To explore the advantages, we triangulated data using mixed methods design, aiming to show quantitative advantages of the VAs in their perceived rapport and eye movement, and to describe the qualitative advantages of the VAs in their disclosed mental symptoms during the interview. A total of 55 Japanese university students participated in comprehensive mental health interviews conducted by the VA and RE. Findings show that participants perceived rapport and moved their right eyes more often, along with disclosing numerous mental symptoms, with the RE than the VA. However, they disclosed more sex-related symptoms to the VA than the RE. The VA can be used most practically in sex-related health fields. The anonymity conditions in the VA setting might be relevant to patients' self-disclosure of sex-related topics.

Keywords: virtual agent, clinical interview, eye movement, threshold model of social influence, rapport, self-disclosure of mental symptoms

[1] Abbreviations
GAF: Global Assessment of Functioning; RE: Real Expert; RH: Real Human; VA: Virtual Agent

Advantages of Virtual Agents over Clinical Psychologists during Comprehensive Mental Health Interviews Using a Mixed Methods Design

1. Introduction

The virtualization of the mental health field began half a century ago (Weizenbaum, 1966); this has entailed the incorporation of virtual reality, a computer technology that creates an artificial environment into which a user's physical presence is simulated, into mental health settings. This has already exhibited positive effects for the treatment of people with anxiety disorders (Kim et al., 2017; Morina, Ijntema, Meyerbröker, & Emmelkamp, 2015). Furthermore, in a related development, Virtual-Agent (VA) technology has also shown positive treatment effects for people with speech disorders (van Vuuren & Cherney, 2014). VAs also have advantages over Real Experts (REs) in facilitating participants' self-disclosure during mental health interviews (DeVault et al., 2014; Kissinger et al., 1999; Kobak et al., 1997; Macalino, Celentano, Latkin, Strathdee, & Vlahov, 2002); therefore, clarification of the advantages is worthwhile in these fields. A VA can be used anytime and anywhere, with little cost (Kazdin & Blase, 2011); thus, using a VA could reduce the cost of psychological assessments (Kobak et al., 1997). Further, a VA also examines physiological data, such as facial movements, through analysis of videos and pictures (Rizzo et al., 2016); these data can afford a detailed analysis of participants' negative emotional expressions during an interview (Lucas, Gratch, King, & Morency, 2014). In summary, a clarification of the VAs' advantages over REs can promote both practical applications of VAs and a detailed analysis of human responses in the mental health

field (Rizzo et al., 2016). Our study compares comprehensive mental health interviews conducted by the audio-visual VA and clinical psychologist as an RE and explores the advantages of the VA over RE through a triangulation mixed methods design (Doyle, Brady, & Byrne, 2009).

1.1. Threshold model of social influence from Virtual Agents

The rationale of our study is based on the threshold model of social influence from VAs, where participants' agency belief and VA's behavioral realism evoke their social response to the VA (Blascovich, 2002). Their agency belief is the extent to which they believe the VA to be a representation of the Real Human (RH)(Lucas et al., 2014). For example, those who believe that they play card games with human operators have higher agency belief than those who believe that they play with computers, even though both of them play with the same computer (Blascovich & Beall, 2010). The threshold model predicts that their high agency belief evokes their social response (Blascovich et al., 2002). Actually, those who believed that they are playing with human operators followed the social norms more frequently than did those who believed that they are playing with computers (Blascovich et al., 2002). Similarly, participants who believed that they were monitored by human operators showed worse performance in novel tasks than did those who believed that they were monitored by computers (Hoyt, Blascovich, & Swinth, 2003). Their worse performance was also common when they were monitored by RH than when they were alone (Zajonc, 1965). These findings indicate that participants' agency belief produces imaginary social influences from the VA and evokes their actual social response to the VA.

VA's behavioral realism also evokes social responses to VAs (Blascovich, 2002). Behavioral realism is the degree to which the VA behaves like the RH. The VA with facial movements has high behavioral realism than the VA without them (von der Pütten, Krämer, Gratch, & Kang, 2010). Actually, participants kept more interpersonal distance from VAs with a gaze feature than the VAs without the gaze feature, even though they knew that the VAs were computer systems (Bailenson, Blascovich, Beall, & Loomis, 2003). Similarly, they also gave good oral presentations to positive VAs (focusing on their presentation) but not to negative VAs (not focusing on their presentation) because VA's positive (negative) listening behaviors promoted (inhibited) their oral presentations (Pertaub, Slater, & Barker, 2002). These findings suggest that VA's high behavioral realism produces imaginary social influence and evokes their actual social response to the VA. The threshold model indicates that participants' high agency belief and VA's high behavioral realism has great social influence on participants and evokes their social response to the VA (Blascovich, 2002).

1.2. Threshold model of social influence from Virtual Agents' interview

Although high social influences by the VA is beneficial in several fields (Blascovich & Beall, 2010), low social influences by the VA is beneficial in interview setting (Rizzo et al., 2016): low social influences from the VA allow participants to be anonymous (DeVault et al., 2014), which has been found to decrease participants' social desirability (Richman, Kiesler, Weisband, & Drasgow, 1999) and increase their self-disclosure (Tidwell & Walther, 2002). Hence, VA's low behavioral realism is beneficial in the interview setting. Actually, voice-only interviews promoted interviewees' self-

disclosure and facial expressions more frequently than face-and-voice interviews (Bailenson, Yee, Merget, & Schroeder, 2006). Similarly, participants' low agency belief is beneficial in interview setting. Compared to participants who believed that they were being interviewed by a human operator, participants who believed that they were being interviewed by a computer experienced lower levels of fear of self-disclosure and more frequently displayed facial expressions relating to sadness (Lucas et al., 2014). These findings indicate that low behavioral realism and low agency belief evoked participants' self-disclosure and facial expressions, although their self-disclosure also occurred under the combination of high behavioral realism and low agency belief conditions (Kang & Gratch, 2014; von der Pütten et al., 2010).

1.3. Threshold model of social influence from Virtual Agents' mental health interview

According to the threshold model (Blascovich, 2002), the RE mental health interview has high behavioral realism (RE is a RH) and evokes high agency belief (Participants believe RE as an RH), whereas VA mental health interview has low behavioral realism and evokes low agency belief. Low behavioral realism and low agency belief were positively linked with self-disclosure (Bailenson et al., 2006; Lucas et al., 2014), and so VA interviews have advantages over RE in terms of interviewees' self-disclosure. For example, binge drug use, prostitution, and engaging in unprotected sexual intercourse are more frequently reported to text-based VAs than to face-to-face-based REs (Kissinger et al., 1999; Macalino et al., 2002). Another study found that participants are more likely to disclose alcohol abuse and mood symptoms to audio-

based VAs than to telephone-based REs (Kobak et al., 1997). Further, participants experience more rapport with a VA (operated by a human operator but they thought it was a computer) than a semiprofessional RE (DeVault et al., 2014; Rizzo et al., 2016). These findings supported the threshold model and confirmed the advantages of VA over RE in mental health interview.

However, the combination of RE's comprehensive mental health interview session (Kobak et al., 1997) and audio-visual VA interview session (Robb et al., 2015) was still rare such that the advantages of VAs over REs during comprehensive mental health interview were unclear. Clarification of these advantages contribute to expanding of the scope of the threshold model (Blascovich, 2002) to RE and promoting practical application of VA in comprehensive mental health interviews. The research question of our study, then, is how does the audio-visual VA outperform the RE during comprehensive mental health interviews? We used comprehensive mental health interviews with VAs and REs to provide a direct comparison between them. To assess performance during VA and RE interviews, we chose to evaluate interviewees' perceived rapport, negative emotional expression, and self-disclosure of mental health symptoms, all of which were key performance indicators during mental health interviews (Duggan & Parrott, 2001; Elvins & Green, 2008; Marci, Ham, Moran, & Orr, 2007).

According to the threshold model (Blascovich, 2002; Lucas et al., 2014), previous direct comparisons between VAs and RHs (Bailenson et al., 2006; Lucas et al., 2014), and the comparisons between VAs and REs (DeVault et al., 2014; Kissinger et

al., 1999; Kobak et al., 1997; Macalino et al., 2002; Rizzo et al., 2016), we have two quantitative hypotheses. Hypothesis 1: Participants perceive a higher level of rapport with the VA than the RE. Hypothesis 2: Participants provide a higher level of negative emotional expression to the VA than the RE. The VA condition could be positively linked with the degree of self-disclosures (Kissinger et al., 1999; Kobak et al., 1997; Macalino et al., 2002), but interview items were inconsistent in previous studies so that integration of these findings were difficult. Instead, we qualitatively explore advantages of the VA over the RE in their self-disclosure of specific symptoms. Hypothesis 3: Participants provide a higher level of self-disclosure in case of specific mental symptoms to the VA than the RE. We quantitatively evaluated their perceived rapport and negative emotional expression by assessing self-reported questionnaires and eye movements, respectively. We also qualitatively evaluated the participants' self-disclosure of mental health symptoms by assessing remarks made during the interviews.

1.4. Evaluation of Performance in Mental Health Interview

To evaluate perceived rapport, participants answered a self-reported questionnaire (Elvins & Green, 2008) after they had completed both VA and RE sessions. To assess negative emotional expression, we recorded participants' eye movements during the interviews, as eye movements have previously been linked to negative emotional expression (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001); for instance, widened and narrowed eyes are considered to represent fear and disgust, respectively (D. H. Lee, Mirza, Flanagan, & Anderson, 2014). Consequently, participants' eye movements were used as indices of negative emotional expressions

during the interviews. Although the Asian population especially emphasizes eye movement while showing emotion (Yuki, Maddux, & Masuda, 2007), we also sampled other facial movements, including those of the mouth, brow, nose, and jaw, because these parts have also been found to reflect emotion (Ekman, 2003). For the evaluation of self-disclosure of mental symptoms, participants' remarks regarding their mental symptoms were used. The diagnosis of their symptoms was also used as a cut-off point for their self-disclosure of specific mental symptoms.

1.5. Aims of the Present Study

The present study utilized audio-visual VA (Lucas et al., 2014; Rizzo et al., 2016; Robb et al., 2015) and compared performance, including interviewees' perceived rapport, negative emotional expression, and self-disclosure of mental symptoms, between VA and RE during comprehensive mental health interview sessions. To elucidate the advantages of the VA over the RE, we triangulated data using mixed methods design (Doyle et al., 2009). The design enabled us to measure and analyze both quantitative (perceived rapport and negative emotional expression) and qualitative (self-disclosure of mental symptom) data. Both types of data were simultaneously sampled during comprehensive mental health interviews using the VA and the RE. Our quantitative data tested the threshold model in perceived rapport and negative emotional expression, whereas our qualitative data explored the advantage of VA in specific mental symptoms. Integration of quantitative and qualitative data could validate our results from two different perspectives.

2. Methods

174 **2.1. Participants**

175 A priori power analysis required 98 participants to detect medium effect size
176 (*Cohen's F* = 0.25) with 80 % (power) and alpha at 0.05 using mixed ANOVA between
177 factors with a repeated measure (correlations between repeated measures were set
178 as .05) (Faul, Erdfelder, Buchner, & Lang, 2009). However, we only sampled 57
179 university students from a Japanese national university because of our limited research
180 resources. These students were recruited by asking a university professor to make an
181 announcement during a psychology class, and also through snowball sampling that
182 involved identifying students' friends through referrals. Of the 57 students, two were
183 excluded because one refused to participate and the other did not attend our laboratory;
184 consequently, our final sample comprised 55 students. All of the participants provided
185 written informed consent and received a gift card (1500 yen, around 12 Euro) in return
186 for their participation. Of the 55 students, 30 were female and 25 were male, and their
187 average age was 22.92 years (95% Confidence Interval [CI]: 22.18, 23.68); further, their
188 mean score of the Global Assessment of Functioning (GAF) was 70.25 (95% CI: 68.16,
189 72.35); hence, the majority of the participants belonged to a non-clinical sample. The
190 GAF was evaluated in the RE sessions, and so the GAF was not balanced during
191 randomization, although the correlations of GAF with order of sessions (RE first or VA
192 first) and participants' gender (male or female) were lower than a medium effect size
193 (Cohen, 1988). All participants were native Japanese speakers and were not regular
194 visitors to mental hospitals or counseling centers. All participants were familiar with the
195 spoken-dialog system in their mobile phone; this meant that they would be familiar with

the system used in the VA settings. However, as they were not regular visitors to mental hospitals, they might have been unfamiliar with mental health interviews as used in the RE settings.

2.2. Experimental Design

All participants were randomly assigned to either the VA first or RE first group; 28 participants were interviewed in the VA scenario first and then in the RE scenario, while 27 participants participated in the RE scenario first, followed by the VA scenario. During randomization, participants' genders, ages, and departments were balanced. For both the VA and RE, a comprehensive, one-on-one mental health interview was conducted with each participant in spoken language. Table 1 shows the durations of the VA and RE sessions; the RE time involved screening questions and utilized strict diagnostic criteria (Table 1), and so it took more time than did the VA ($F = 58.48$; $df = 1, 51$; *Cohen's F* = 1.071, *generalized* $\eta^2 = .33$), even though the order effect of sessions (RE first or VA first) and gender (male or female) were controlled. Furthermore, both VA and RE used preliminary questions to check for specific symptoms. If the participants confirmed the symptoms, both the VA and the RE inquired about the symptoms in detail so that the duration of their interview was lengthened. If not, they did not ask about the symptoms in detail, so their interview was shorter. The durations of interviews in both settings were changed based on participant responses; hence, the durations were not controlled. After the participants completed their VA and RE interviews, they completed the rapport questionnaire and proceeded to a debriefing session. During this session, they received feedback from the RE concerning their

mental symptoms; further, the RE answered questions and recommended treatment for some participants.

2.2.1. The virtual agent. VA employed a spoken-dialog system (Figure 1A) that uses Julius 4.4.2 (A. Lee & Kawahara, 2009), a Japanese speech-recognition system. It has an inbuilt web camera that takes pictures every 500 milliseconds (refer to (Yokotani, 2016) for additional mechanical details concerning this system). The VA administered the Japanese version (Otsubo et al., 2005) of the Mini-International Neuropsychiatric Interview 5.0 (Sheehan et al., 1997). Table 1 shows the question formats. The VA did not obtain the patients' past mental history, but mainly sampled current symptoms, and did not consider the effects of their physical conditions and other mental disorders on suspected disorders.

2.2.2. The virtual agent scenario. Participants conversed with the VA through a microphone and used the computer mouse to advance the conversation (Figure 1B). The applied condition included both training and experimental sessions. During the training session, one experimenter was in the experimental room with the participant, and during this session, he/she adjusted the height of the microphone to suit the participant and instructed the participant in regard to how he/she should talk with the VA (Table 1). After the training session, the experimenter left the room and the VA appeared on the computer monitor. During this experimental session, participants talked alone with the VA, which conducted the comprehensive mental health interview. The experimenter did not return to the session until the participant rang a bell.

2.2.3. The Real Expert. The RE in this study was a male Japanese clinical

psychologist with over 10 years' experience in the mental health field. He also had experience conducting psychological treatment sessions for the inmates of a Japanese prison, as well as mental evaluations for the accused in a Japanese court (Yokotani & Tamura, 2015).

2.2.4. The Real Expert scenario. Participants conversed with the RE in another experimental room (Figure 1C) and during the conversation, their facial expressions were video recorded. The expert administered the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision Axis I disorders, Non-patient edition (First, Spitzer, Gibbon, & Williams, 1997), using the Japanese version (First et al., 2010). Table 1 shows the format of the questions. The RE obtained information about participants' educational, work, and medical history. The RE also inquired about demographic data and social functioning and about symptoms from both current and lifetime perspectives. The RE also considered the effects of their physical condition and other mental disorders on their suspected disorders (Table 1).

2.3. Quantitative Measures and Analysis

2.3.1. Perceived rapport. A previous study recommended assessment of perceived rapport using participants' responses on a self-reported questionnaire (Elvins & Green, 2008). As such, we used a self-reported questionnaire concerning rapport (Kakii, 1997) that comprised six items using a five-point scale. The first two questions pertained to informational communication (e.g., item 1: I felt that what I wanted to say was transmitted to counselor); the next two questions pertained to emotional

communication (e.g., item 4: I felt that the counselor understood my feelings); and the last two questions pertained to trust (e.g., item 6: I felt that the counselor created a warm atmosphere). Participants were asked to respond to this questionnaire, which concerned both the VA and RE, after they had completed both interviews. The mean scores for these subcategories were then used for analysis.

2.3.2. Negative emotional expression: Eye movements. All videos recorded during the RE condition (1280 * 720 pixels, 29.9 frames per second) were converted into a series of pictures that represented one image for every 500 milliseconds of video. For the VA, the pictures taken during the experimental session (800 * 600 pixels) were collected, but the pictures recorded during the training session were discarded because some may have included instances when the experimenter was communicating with the participant. In total, 363,718 pictures were analyzed.

To determine facial landmarks, we used OpenCV and dlib (King, 2009) and, as a result, 68 landmarks for each picture were identified (Figure 1D). The landmarks for the eyes were the right eyes (Marks 37–42) and the left eyes (Marks 43–48). To adjust participants' facial size and rotation, all facial landmarks were transformed to one averaged face picture (320 * 400 pixels) (Langlois & Roggman, 1990) using the affine formula.

We calculated absolute differences in landmarks between each picture and the previous picture (the one that had been taken 500 milliseconds previously). When the landmarks between the two pictures differed in regard to the X axis, we scored the difference as horizontal movement. $\frac{1}{(k2-k1+1)(N-1)}\sum_{k=k1}^{k2}\sum_{n=1}^{N-1}|X_{k,n+1}-X_{k,n}|$;

where $X_{k, n}$ is the x coordinate at time n at position k; N indicates the final time in the picture during the interview; k1 and k2 indicate the start and end positions, respectively. In the case of the right eye area, k1 is position 37 and k2 is position 42. Similarly, when the landmarks differed in relation to the Y axis, we scored the difference as vertical movement. High movement scores indicated a high frequency and wide variety of movements. Similarly, we also calculated movements of the mouth (inner lip: marks 61–68; outer lip: marks 49–60), nose (nasal cavity: marks 28–31; ridge of nose: marks 32–36), eyebrows (right eyebrow: marks 18–22; left eyebrow: marks 23–27), and jaw (marks 1–17), and the averages of these movements during the RE and VA interviews were used for analysis (Table 2).

2.3.3. Quantitative analysis. To compare participants' perceived rapport between the VA and RE scenarios, we utilized a 2 (VA first or RE first) x 2 (male or female) x 2 (perceived rapport) mixed ANOVA with perceived rapport as the repeated measure. The order of sessions and participants' genders were also controlled. Similarly, to compare participants' eye movement (negative emotional expression) between the VA and RE scenarios, we used a 2 (VA first or RE first) x 2 (male or female) x 2 (eye movement) mixed ANOVA with eye movement as the repeated measure. To decrease the risk of Type 1 errors strictly, we adjusted the *p* value of ANOVAs through the Bonferroni method (Armstrong, 2014). Hence, the adjusted *p* values of .01 in these ANOVAs with perceived rapport and eye movement were .0033 (divided by 3) and .0005 (divided by 18), respectively. Additionally, to clarify the effect size of ANOVAs, we utilized *Cohen's F* (Cohen, 1988) and generalized η^2 (Olejnik &

Algina, 2003). *Cohen's F* was calculated as $\sqrt{partial\ \eta^2/(1-partial\ \eta^2)}$(Cohen, 1988). For exploratory correlational analysis among variables, a medium effect size correlation (.3) was used for a cut-off score (Cohen, 1988) and we did not use *p* values, because adjustment and interpretation of *p* values among many exploratory correlations is controversial (Armstrong, 2014).

2.4. Qualitative Measures and Analysis

2.4.1. Self-disclosure of Mental Symptoms. The VA asked participants about mental symptoms and diagnosed mental disorders based on their disclosed symptoms (Sheehan et al., 1997); however, VAs cannot describe developments in mental disorders (i.e., when a disorder started, ended, and restarted), and it occasionally abbreviates the history of disorders (e.g., a current episode of hypomania would abbreviate questions concerning past episodes of hypomania). Therefore, it merely diagnoses whether a disorder is present or not, without providing the length of time for which it has persisted. Furthermore, the length of the current episode is arbitrary. While one disorder may have presented only in the past month, another could have been experienced throughout the participant's lifetime.

In contrast, the RE asked participants about mental symptoms in detail, described the development of mental disorders, and clarified diagnoses based on the development of other disorders. For example, when both social phobia and agoraphobia are simultaneously present and the social phobia explains the agoraphobia, the RE diagnoses social phobia only (First et al., 1997). The RE's question about mental disorder and diagnosis can thus be more comprehensive than that of the VA (Table 1).

Therefore, the RE's diagnosis (cut-off point for participants' self-disclosure in specific mental topics) was adjusted based on the VA's diagnosis. For example, if the duration of current alcohol abuse was one month according to the RE but one year according to the VA, the RE re-diagnosed the current alcohol abuse by adjusting for the one-year duration based on the development of the participant's mental disorders.

2.4.2. Qualitative analysis. To explore the differences of participants' disclosed mental symptoms in the VA and RE settings, four indices were used (Table 3). A true positive indicates that both RE and VA supported the presence of a mental disorder, while a false positive indicates that the VA supported the presence of a disorder but the RE did not. Furthermore, a false negative indicates that the RE supported the presence of a disorder but the VA did not, while a true negative indicates that neither the VA nor the RE supported the presence of a disorder. Remarks regarding the patients' symptoms were also analyzed.

2.5. Ethical Considerations

The current study was approved by an ethics committee of a national university in Japan. Furthermore, all procedures were conducted in accordance with guidelines for studies involving human participants, the ethical standards of the institutional research committee, and the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

3. Results

3.1. Quantitative Findings

3.1.1. Correlations over the medium effect size among participants'

gender, facial movement, perceived rapport, and order of sessions. We explored the correlations among control variables (participants' gender and order of sessions) and dependent variables (facial movement and perceived rapport). We did not find any correlations between the order of the sessions and the other variables. Hence, we describe only the correlations among the other variables.

Participants' gender was correlated with their perceived rapport and facial movements in both the RE and the VA setting. Male participants moved the bridge of their noses in VA settings more often than did their female counterparts (r = .330, N = 54). Male participants perceived less emotional warmth in RE settings than female participants (r = -.325, N = 55). Female participants perceived more warmth with the RE but moved their noses less during VA sessions than did the male participants.

Facial movements in the VA setting were also correlated with perceived rapport with the VA. Horizontal and vertical movements of the participants' jaws in the VA session were also positively correlated with their perceived emotional warmth with the VA (r = .313, N = 54; r = .307, N = 54, respectively). Horizontal movements of the participants' right and left eyes in the VA session were also positively correlated with their perceived emotional warmth toward the VA (r = .360, N = 54; r = .309, N = 54, respectively). Interestingly, the horizontal movement of the participants' jaws and inner lips in the VA settings were negatively correlated with their perceived warmth toward the RE (r = -.309, N = 54) and their perceived trust in the RE (r = -.311, N = 54). These findings indicated positive correlations between participants' facial movements in VA sessions and perceived emotional warmth toward the VA.

Facial movements in the RE sessions were correlated with perceived rapport with the RE. The participants' horizontal movements of outer and inner lips in the RE session were positively correlated with perceived warmth toward the RE (r = .320, N = 54; r = .304, N = 54, respectively). The vertical movements of their outer and inner lips in the RE session were also positively correlated with their perceived warmth toward the RE (r = .310, N = 54; r = .385, N = 54, respectively). The vertical movement of their right eyes in the RE session was also positively correlated with their perceived warmth toward the RE (r = .328, N = 54). These findings indicate positive correlations between facial movements in the RE session and perceived emotional warmth toward the RE.

3.1.2. Comparison of perceived rapport between the Virtual Agent and Real Expert. We compared the perceived rapport between the VA and RE scenarios (Figure 1E) and found that in all three subcategories, the score for the VA condition was significantly lower than that for the RE condition (information: F = 203.55; df = 1, 51; *Cohen's F* = 1.998, *generalized* η^2 = .68, a*djusted p* < .0001; emotion: F = 181.49; df = 1, 51; *Cohen's F* = 1.886, *generalized* η^2 = .59, a*djusted p* < .0001; trust: F = 203.55; df = 1, 51; *Cohen's F* = 1.998, *generalized* η^2 = .68, a*djusted p* < .0001), even though the order of sessions and participants' gender were controlled. Thus, compared to a VA, participants perceive greater rapport with an RE.

3.1.3. Comparison of participants' eye movements between the Virtual Agent and Real Expert. Next, we compared participants' eye movements between the VA and RE sessions. As an example, Figure 1 shows one participant's facial movements during the VA (Figure 1F) and RE scenarios (Figure 1G); these images

indicate that eye movements during the VA scenario were narrower than those in the RE scenario. Table 2 presents the findings for the comparison of participants' facial movements between the VA and RE scenarios. This shows that there were significantly more vertical and horizontal movements of the right eye in the RE scenario than in the VA scenario, even though we strictly adjusted *p* values. On the other hand, there were no significant differences in the participants' left eye movements between the VA and RE scenarios. Further, the *Cohen's F* and *generalized* η^2 also indicated that differences between the vertical movements of the participants' right eyes were higher than large effect sizes.

Interestingly, the participants' jaws were found to more often move vertically in the RE scenario than did those in the VA scenario. These movements could indicate that the participants talked more in the former scenario than in the latter. Furthermore, we also found different lateralization between the RE and VA scenarios. In the RE scenario, participants' right eyes moved more than did their left eyes (vertical movement: $F = 40.25$; $df = 1, 50$; *Cohen's F* = .897, *generalized* $\eta^2 = .09$), while in the VA condition, their left eyes moved more than did their right eyes (horizontal movement: $F = 15.42$; $df = 1, 50$; *Cohen's F* = 0.555, *generalized* $\eta^2 = .022$). For both, *Cohen's F* was higher than a large effect size.

3.1. Qualitative Findings

3.1.1. Comparison of self-disclosure for each mental topic between the Virtual Agent and Real Expert.

Table 3 shows the comparison of self-disclosure for each mental topic between the VA and RE. During the topic of mood, the participants

disclosed their past major depressive symptoms more often to the RE than to the VA. Similarly, during the topic of anxiety, they also disclosed their anxious symptoms, especially social phobic symptoms, more often to the RE than to the VA. A comparison of participants' comments regarding these symptoms clarified that the participants disclosed their hypomanic, depressive, and anxiety-related concerns to the RE in detail, but not to the VA.

During the topic of alcohol, the participants disclosed the same degree of alcohol-related symptoms (i.e., current alcohol dependence and abuse) to the VA and to the RE. However, detailed analysis of participants' alcohol-related remarks showed several deficits in the VA's alcohol-related symptoms. Three false negative cases, that is, cases where the VA did not diagnose alcohol abuse, had actually experienced alcohol-related blackout episodes during the last year, which is sufficient to reach the threshold of alcohol abuse (First et al., 1997). Furthermore, for the five false positive cases, that is, where the VA supported the existence of an alcohol-abuse condition, four had not reached the threshold of alcohol abuse during the last year, while the other had never reached it in their lifetime. A possible explanation for these discrepancies is that the VA allowed participants to define their mental problems subjectively; therefore, participants' subjectivity might have contaminated the objective symptoms (e.g., for some individuals, an alcohol-related blackout is not a mental problem and, consequently, they give a negative response when asked if they have ever experienced an alcohol-related problem).

In addition, the topic of eating disorders returned four false positives. In all

cases, detailed analyses showed that participants disclosed severe symptoms to the female VA but not to the male RE. For example, one female participant with anorexia reported her menstruation had stopped to the VA but did not to the RE. One male participant with anorexia reported his weight as fewer than 51 kg to the VA, but he reported his weight as more than 52 kg to the RE. Another female participant with bulimia reported that she engages in binge eating more than twice per week to the VA, but the same participant reported that her binge eating occurs once in six months to the RE. The other female participant with bulimia reported binge eating to the VA, but did not to the RE. A possible explanation is that these four cases found it easier to disclose sex-related topics, such as their menstruation and body shape, to the female VA than to the male RE.

4. Discussion

4.1. Contribution of Quantitative Findings to Threshold Model of Social Influence from Virtual Agent in Mental Health Interview

Our quantitative findings did not support the threshold model of social influence from VAs (Blascovich, 2002). In contrast to our quantitative hypotheses, participants perceived emotional warmth and moved their right eyes more often with the RE than with the VA. Their perceived emotional warmth toward the RE was also positively linked with the horizontal movement of their right eyes during the RE session. Similarly, their perceived emotional warmth toward the VA was positively correlated with the vertical movement of their right eyes during the VA session. These findings indicate positive associations among the RE session, their perceived emotional warmth,

and right eye movement. Inconsistencies between previous and current findings might be from the differences between RE and RH, which is outside the scope of the threshold model because the model mainly focused on the differences between VA and RH (Blascovich et al., 2002; Blascovich & Beall, 2010; Hoyt et al., 2003). Many studies reported that human responses are different to RE and novice RH interviews. People who believed the interviewer as an authority followed the interviewer's guidance more frequently than those who believed the interviewer as a novice interviewer (Blass, 1999). They also perceived rapport more often with the expert interviewer than with the novice interviewer (Mallinckrodt & Nelson, 1991). These findings indicate that human responses to current RE interviews could be different from those to previous RH interviews (Bailenson et al., 2006).

The RE interview session in our study might exceed the high behavioral realism and high agency belief conditions: the session might have hyper-high behavioral realism (interviewer behaves like RE) and evoke hyper-high agency belief (they believe the interviewer as a representation of RE). The hyper-high behavioral realism could especially be linked with perceived emotional warmth and movement of their right eyes with the RE, because the hyper-high behavioral realism was positively linked with their perceived rapport (Ramseyer & Tschacher, 2011) and negative emotional expression (Marci et al., 2007). On the other hand, in the low behavioral realism (VA did not behave like RH) and low agency belief conditions (participants did not believe the VA as a representation of RH), the number of times that they thought that the VA interviewer understand their intention immediately might have been lower than that for

the RE interviewer (Gratch, Wang, Gerten, Fast, & Duffy, 2007), and so that they did not feel enough warmth toward the VA and did not move their right eyes as much in the VA session. The hyper-high behavioral realism matched with expert-novice studies (Mallinckrodt & Nelson, 1991) and explained high perceived rapport in previous VA condition and low rapport in previous semi-RE condition (DeVault et al., 2014; Rizzo et al., 2016). VA operated by a human operator could result in the hyper-high behavioral realism, such that they perceived rapport more often with VAs operated by human than with VAs operated by a computer system. The previous semi-professional RE condition might not show hyper-high behavioral realism like current RE condition because training duration of semi-professional RE was limited (Mallinckrodt & Nelson, 1991), and so participants might not perceive rapport with the semi-professional RE like the current RE (Mallinckrodt & Nelson, 1991). Improvement of VA's nonverbal behaviors in recent studies imply that their VA might exceed the high behavioral realism and demonstrate hyper-high behavioral realism (Kang & Gratch, 2014; von der Pütten et al., 2010). The threshold model needs to include the hyper-high behavioral realism condition because the advanced VA could exceed novice RH in terms of behavioral realism (Rizzo et al., 2016).

Our findings also indicate that participants' responses to the VA were different from their responses to the RE. For example, jaw movements in VA sessions were positively linked with perceived emotional warmth toward the VA, but in RE sessions they were not linked with perceived emotional warmth toward the RE. Further, participants' right eyes moved vertically more than their left eyes did in RE sessions,

whereas their left eyes moved horizontally more than their right eyes did in VA sessions. In line with previous comparisons between VAs and RHs (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), these findings indicate that the links between facial movement and perceived emotional warmth in VA sessions are different from those in RE sessions. Further, eye movements were mainly linked with negative emotions (Baron-Cohen et al., 2001; D. H. Lee et al., 2014; Yuki et al., 2007). Hence, narrow movement of right eyes in the VA sessions could indicate that participants did not feel negative emotions in the session. A possible interpretation of the narrower eye movements in VA sessions is that participants perceived less warmth toward the VA than toward the RE, but at the same time, they might feel less negative emotion in the VA session than the RE session. This interpretation implies that their feelings in the VA session and in the RE session might not be in the same dimensions. Although the threshold model explains individuals' consecutive experiences of VA and RE sessions (Blascovich & Beall, 2010), their experiences in VA and RE sessions might not be connected to each other.

4.2. Contribution of Qualitative Findings to Threshold Model of Social Influence from Virtual Agent in Mental Health Interview

Qualitative findings were both positive and negative regarding the threshold model (Blascovich, 2002). From the positive perspective, our study found that the VA had an advantage over the RE regarding self-disclosure of eating disorder symptoms. All four participants with eating disorders reported more severe symptoms about sex-related topics to the VA than the RE. These were consistent with previous findings (Kissinger

et al., 1999; Macalino et al., 2002). People with eating disorders have been found to possess a fear of evaluation by a male individual regarding their physical attractiveness (Siever, 1994); this may have hindered them from giving honest responses regarding their eating habits and weight during their interview with the male RE. In contrast, a female VA setting, as a computer system, could provide more anonymity to participants than the RE setting and might decrease such a fear of evaluation and promote honest disclosure (Lucas et al., 2014).

The main factor that decreases participants' fear of evaluation in the VA session may be the setting's anonymity. VA involves computer-mediated communication that allows participants to enjoy anonymity (Kiesler, Siegel, & McGuire, 1984), which has been found to decrease participants' social desirability (Richman et al., 1999) and increase their self-disclosure (Tidwell & Walther, 2002). Thus, the VA might provide anonymity and promote self-disclosure about sex-related topics such as menstruation and body weight. This interpretation supports a previous study that found an advantage of text-based VAs over face-to-face-based REs regarding interviewees' self-disclosure of sexual behaviors (Kissinger et al., 1999; Macalino et al., 2002). As predicted by the threshold model, low behavioral realism and low agency belief conditions could be an advantage of the VA over the RE regarding participants' self-disclosure of sex-related symptoms.

On the other hand, participants disclosed mood and anxiety symptoms to the RE more often than the VA. Although these findings are inconsistent with previous findings, which have shown the existence of a comparable amount of self-disclosure

regarding mood and anxiety symptoms between VAs and REs (Kobak et al., 1997), this inconsistency might stem from the visual modality: the VA and RE methods applied in this study used visual modality as a communication avenue, but previous VA and RE did not. Our VA's visual modality involved the use of a camera; however, this may have weakened the VA's anonymity, as interviewees may have noticed that they were being monitored by the camera. Furthermore, RE's visual modality might have improved the RE's nonverbal synchronized behaviors with interviewees, because the RE could verify his synchronization with them by observing their body language. RE's nonverbal synchronization behavior was positively linked with patients' perceived rapport (Ramseyer & Tschacher, 2011) and self-disclosure (Duggan & Parrott, 2001). Hence, the visual modality applied in our study might have demoted VA's anonymity and participants' self-disclosure of mood/anxiety symptoms, while promoting RE's perceived rapport and their self-disclosure.

Further, participants disclosed the same degree of alcohol-related symptoms to the VA and to the RE, but their self-disclosure in VA scenario was based on their subjective evaluation rather than on an expert's objective evaluation. Although a previous study indicated an advantage for the VA regarding alcohol disorders (Kobak et al., 1997), our results indicate that alcohol-related symptoms disclosed in the VA scenario might include participants' amateur judgment and might not represent actual alcohol-related symptoms (First et al., 1997). Mental health interviews of VA and RE during mood, anxiety, and alcohol topics indicate disadvantages of the VA relative to the RE in their self-disclosure of mood, anxiety, and alcohol-related symptoms.

4.3. Integration of Quantitative and Qualitative Findings in Mental Health Virtual Agents

Our quantitative and qualitative findings both support and contrast the threshold model in mental health interview (Bailenson et al., 2006; Blascovich et al., 2002; Kang & Gratch, 2014; Lucas et al., 2014; von der Pütten et al., 2010). Integration of these findings indicates that the low agency belief and low behavioral realism are linked with participants' self-disclosure patterns about the sex-related topics, while hyper-high agency belief (Participant believe interviewer as a representation of RE) and hyper-high behavioral realism (interviewer behave like RE) are linked with the perceived rapport, negative emotional expression, and the self-disclosure of topics related to mood/anxiety. For the sex-related topic, they might disclose their secrets when their perceived social pressures were lower than their own threshold. To decrease social pressures, low agency belief is beneficial. Non-human guidance, such as a "machine automatically interviewing you" might lower their agency belief (Lucas et al., 2014). Similarly, low behavioral realism of VA is also beneficial. Text-only VA without voice and face could lower behavioral realism and encourage their self-disclosure of sex-related topics (Kissinger et al., 1999; Macalino et al., 2002). According to the threshold model of social influence, low agency belief and low behavioral realism lower their perceived social pressure and enhance their self-disclosure of sex-related topics.

On the other hand, for the mood and anxiety topics, interviewees might disclose their secrets when the number of times their intention was understood immediately were higher than that of their own threshold (Gratch et al., 2007). To enhance the number, the

hyper-high behavioral realism is helpful. Interviewer's nonverbal synchronization features allows interviewees to believe that the interviewer understood them immediately (Chartrand & Bargh, 1999). The nonverbal synchronization features were frequently used in RE interview sessions and correlated with their perceived emotional warmth toward the interviewer, negative emotional expression during the interview, and self-disclosure of their symptoms (Duggan & Parrott, 2001; Marci et al., 2007; Ramseyer & Tschacher, 2011). The addition of such features to VAs could improve their perceived warmth toward the VA, negative emotional expression during interviews, and self-disclosure of such symptoms (Kang & Gratch, 2014; von der Pütten et al., 2010).VA's highly-communicative features could be associated with their self-disclosure of topics related to mood and anxiety, whereas VA's non-human features could be associated with their self-disclosure of sex-related topics.

4.4. Limitations of our study

Our study had limitations in participant sampling, experimental condition, and diagnostic criteria. First, the number of participants was not high enough to warrant a medium effect size. The power of medium effect size of *Cohen's F* (0.25) in our settings ($N = 55$, the number of groups = 2, the number of measurements = 2, correlation among the repeated measures = 0.5) was suggested to be 0.55 and approached the level of randomness, although the power of large effect size of *Cohen's F* (0.40) in our settings indicated enough power (0.91) (Faul et al., 2009). Snowball sampling also biased the participants, and GAF scores indicated that participants were mainly non-clinical samples. Their medical history indicated that they were unfamiliar

with mental health interviews, and so the ecological validity of the RE setting might not be good from the perspective of clinical patients (Kobak et al., 1997).In other words, our study is still an analog study in mental health settings. More clinical participants are required in future studies. Second, our experimental condition included only one female VA and one male RE; hence, the gender effect of interviewers might contaminate the findings, as interviewers' genders might be a confounder of duration and contents of interviews. Furthermore, VA features were not manipulated, and so it is unclear which features affected participants' responses. Future research needs to involve multiple, diverse VAs and REs. Third, our diagnostic criteria included neither physical examinations nor interviews with third parties such as participants' parents. The lack of physical examinations did not completely exclude direct effects of their physical conditions on their mental disorders. Furthermore, the lack of third-party interviews might increase the risk of wrong diagnosis. Some participants with mental disorders might disclose their symptoms neither to an RE nor to a VA. Future study needs to involve physical data and third-party interviews.

5. Conclusion

Despite these limitations, to our knowledge, the current study is the first to clarify advantages of the audio-visual VA over an RE during a comprehensive audio-visual mental health interview. Although the VA has several disadvantages in terms of rapport building, negative emotional expression, and objective symptom evaluation, it showed advantages for participants regarding sex-related topics. The VA's provision of anonymity might help such people to disclose their symptoms without fear of

636 evaluation; hence, VAs could be particularly useful in sex-related fields, such as in
637 sexual addiction clinics. Further, the implementation of VAs in mental health interviews
638 improves data sampling, analysis, and synthesis (Yarkoni, 2012); actually, our findings
639 show positive links between participants' right eye movements during interviews and
640 their perceived warmth toward the interviewer. Therefore, application of the VA in sex-
641 related mental health interviews could improve patients' self-disclosure of sexual
642 symptoms and detailed analysis of their responses during the interviews.

643 **6. References**

644 Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and*
645 *Physiological Optics*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131

646 Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal
647 distance in immersive virtual environments. *Personality and Social Psychology*
648 *Bulletin*, *29*(7), 819–833. https://doi.org/10.1177/0146167203029007002

649 Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral
650 realism and form realism of real-time avatar faces on verbal disclosure,
651 nonverbal disclosure, emotion recognition, and copresence in dyadic interaction.
652 *Presence: Teleoperators and Virtual Environments*, *15*(4), 359–372.
653 https://doi.org/10.1162/pres.15.4.359

654 Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading
655 the mind in the eyes" test, revised version: A study with normal adults, and
656 adults with Asperger syndrome or high-functioning autism. *Journal of Child*
657 *Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1111/1469-
658 7610.00715

659 Blascovich, J. (2002). A theoretical model of social influence for increasing the utility
660 of collaborative virtual environments. In *Proceedings of the 4th International*
661 *Conference on Collaborative Virtual Environments* (pp. 25–30). New York, NY,
662 USA: ACM. https://doi.org/10.1145/571878.571883
663 Blascovich, J., & Beall, A. C. (2010). Digital immersive virtual environments and
664 instructional computing. *Educational Psychology Review*, *22*(1), 57–69.
665 https://doi.org/10.1007/s10648-010-9120-0
666 Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N.
667 (2002). Immersive virtual environment technology as a methodological tool for
668 social psychology. *Psychological Inquiry*, *13*(2), 103–124.
669 https://doi.org/10.1207/S15327965PLI1302_01
670 Blass, T. (1999). *Obedience to Authority: Current Perspectives on the Milgram*
671 *Paradigm*. Psychology Press.
672 Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior
673 link and social interaction. *Journal of Personality and Social Psychology*, *76*(6),
674 893–910. https://doi.org/10.1037/0022-3514.76.6.893
675 Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Edition).
676 New York, NY: Routledge.
677 DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., … Morency, L.-P.
678 (2014). SimSensei Kiosk: A virtual human interviewer for healthcare decision
679 support. In *Proceedings of the 2014 International Conference on Autonomous*
680 *Agents and Multi-agent Systems* (pp. 1061–1068). Richland, SC: International

Foundation for Autonomous Agents and Multiagent Systems. Retrieved from http://dl.acm.org/citation.cfm?id=2617388.2617415

Doyle, L., Brady, A.-M., & Byrne, G. (2009). An overview of mixed methods research. *Journal of Research in Nursing*, *14*(2), 175–185. https://doi.org/10.1177/1744987108093962

Duggan, P., & Parrott, L. (2001). Physicians' nonverbal rapport building and patients' talk about the subjective component of illness. *Human Communication Research*, *27*(2), 299–311. https://doi.org/10.1111/j.1468-2958.2001.tb00783.x

Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, *1000*(1), 205–221.

Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, *28*(7), 1167–1187. https://doi.org/10.1016/j.cpr.2008.04.002

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders*. Washington, DC: American Psychiatric Publishing, Inc.

First, M. B., Spitzer, R. L., Miriam, G., Williams, J. B. W., Kitamura, T., Okano, T., & Takahashi, S. (2010). *SeishinkashindanmensetsumanualSCID: shiyonotebiki/tesutoyoshi* (2nd edition). Nihonhyoronsha. Retrieved from

http://ci.nii.ac.jp/ncid/BB02398496

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In *Intelligent Virtual Agents* (pp. 125–138). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74997-4_12

Hoyt, C. L., Blascovich, J., & Swinth, K. R. (2003). Social inhibition in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, *12*(2), 183–195. https://doi.org/10.1162/105474603321640932

Kakii, T. (1997). Characteristics of multimedia counseling: A study of an interactive TV system. *The Japanese Journal of Psychology*, *68*(1), 9–16. https://doi.org/10.4992/jjpsy.68.9

Kang, S.-H., & Gratch, J. (2014). Exploring users' social responses to computer counseling interviewers' behavior. *Computers in Human Behavior*, *34*, 120–130. https://doi.org/10.1016/j.chb.2014.01.006

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science*, *6*(1), 21–37. https://doi.org/10.1177/1745691610393527

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*(10), 1123–1134. https://doi.org/10.1037/0003-066X.39.10.1123

Kim, H. E., Hong, Y.-J., Kim, M.-K., Jung, Y. H., Kyeong, S., & Kim, J.-J. (2017). Effectiveness of self-training using the mobile-based virtual reality program in patients with social anxiety disorder. *Computers in Human Behavior*, *73*, 614–

619. https://doi.org/10.1016/j.chb.2017.04.017

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, *10*(Jul), 1755–1758.

Kissinger, P., Rice, J., Farley, T., Trim, S., Jewitt, K., Margavio, V., & Martin, D. H. (1999). Application of computer-assisted interviews to sexual behavior research. *American Journal of Epidemiology*, *149*(10), 950–954. https://doi.org/10.1093/oxfordjournals.aje.a009739

Kobak, K. A., Taylor, L. vH, Dottl, S. L., Greist, J. H., Jefferson, J. W., Burroughs, D., … Serlin, R. C. (1997). A computer-administered telephone interview to identify mental disorders. *JAMA*, *278*(11), 905–910. https://doi.org/10.1001/jama.1997.03550110043034

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1*(2), 115–121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x

Lee, A., & Kawahara, T. (2009). Recent development of open-source speech recognition engine Julius. *Proceedings : APSIPA ASC 2009 : Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, 131–137.

Lee, D. H., Mirza, R., Flanagan, J. G., & Anderson, A. K. (2014). Optical origins of opposing facial expression actions. *Psychological Science*, *25*(3), 745–752. https://doi.org/10.1177/0956797613514451

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer:

747 Virtual humans increase willingness to disclose. *Computers in Human Behavior*,
748 *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043
749 Macalino, G. E., Celentano, D. D., Latkin, C., Strathdee, S. A., & Vlahov, D. (2002).
750 Risk behaviors by audio computer-assisted self-interviews among HIV-
751 seropositive and HIV-seronegative injection drug users. *AIDS Education and*
752 *Prevention*, *14*(5), 367–378. https://doi.org/10.1521/aeap.14.6.367.24075
753 Mallinckrodt, B., & Nelson, M. L. (1991). Counselor training level and the formation of
754 the psychotherapeutic working alliance. *Journal of Counseling Psychology*,
755 *38*(2), 133–138. https://doi.org/10.1037/0022-0167.38.2.133
756 Marci, C. D., Ham, J., Moran, E., & Orr, S. P. (2007). Physiologic correlates of
757 perceived therapist empathy and social-emotional process during psychotherapy.
758 *The Journal of Nervous and Mental Disease*, *195*(2), 103–111.
759 https://doi.org/10.1097/01.nmd.0000253731.71025.fc
760 Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. G. (2015). Can virtual
761 reality exposure therapy gains be generalized to real-life? A meta-analysis of
762 studies applying behavioral assessments. *Behaviour Research and Therapy*, *74*,
763 18–24. https://doi.org/10.1016/j.brat.2015.08.010
764 Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures
765 of effect size for some common research designs. *Psychological Methods*, *8*(4),
766 434–447. https://doi.org/10.1037/1082-989X.8.4.434
767 Otsubo, T., Tanaka, K., Koda, R., Shinoda, J., Sano, N., Tanaka, S., … Kamijima, K.
768 (2005). Reliability and validity of the Japanese version of the Mini-International

Neuropsychiatric Interview. *Psychiatry and Clinical Neurosciences*, *59*(5), 517–526. https://doi.org/10.1111/j.1440-1819.2005.01408.x

Pertaub, D.-P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, *11*(1), 68–78. https://doi.org/10.1162/105474602317343668

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, *79*(3), 284–295. https://doi.org/10.1037/a0023419

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, *84*(5), 754–775. https://doi.org/10.1037/0021-9010.84.5.754

Rizzo, A., Shilling, R., Forbell, E., Scherer, S., Gratch, J., & Morency, L.-P. (2016). Autonomous virtual human agents for healthcare information support and clinical interviewing. In D. D. Luxton (Ed.), *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 53–79). San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-420248-1.00003-9

Robb, A., White, C., Cordar, A., Wendling, A., Lampotang, S., & Lok, B. (2015). A comparison of speaking up behavior during conflict with real and virtual humans. *Computers in Human Behavior*, *52*, 12–21.

https://doi.org/10.1016/j.chb.2015.05.043

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758. https://doi.org/10.1126/science.1082976

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., … Dunbar, G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, *12*(5), 232–241.

Siever, M. D. (1994). Sexual orientation and gender as factors in socioculturally acquired vulnerability to body dissatisfaction and eating disorders. *Journal of Consulting and Clinical Psychology*, *62*(2), 252–260. https://doi.org/10.1037/0022-006X.62.2.252

Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, *28*(3), 317–348. https://doi.org/10.1111/j.1468-2958.2002.tb00811.x

van Vuuren, S., & Cherney, L. R. (2014). A virtual therapist for speech and language therapy. *Intelligent Virtual Agents : --International Workshop, IVA -- Proceedings. IVA (Conference)*, *8637*, 438–448. https://doi.org/10.1007/978-3-319-09767-1_55

von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). “It doesn’t matter what you are!” Explaining social effects of agents and avatars. *Computers*

in Human Behavior, *26*(6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communication of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Yarkoni, T. (2012). Psychoinformatics: New horizons at the interface of the psychological and computing sciences. *Current Directions in Psychological Science*, *21*(6), 391–397. https://doi.org/10.1177/0963721412457362

Yokotani, K. (2016). Automated psychiatric evaluation through spoken dialogue system. Presented at the the 18th Spoken Language Symposium & the Third Natural Language Processing Symposium, Tokyo.

Yokotani, K., & Tamura, K. (2015). Effects of Personalized Feedback Interventions on Drug-Related Reoffending: a Pilot Study. *Prevention Science*, *16*(8), 1169–1176. https://doi.org/10.1007/s11121-015-0571-x

Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, *43*(2), 303–311. https://doi.org/10.1016/j.jesp.2006.02.004

Zajonc, R. B. (1965). Social Facilitation. *Science*, *149*(3681), 269–274.

Acknowledgement

We appreciate Dr. Kurosawa, Tai for insightful feedbacks on our early draft, an anonymous editor for his/her great effort to approach the potential 22 referees for 6 months, and two actual anonymous referees for their intelligent comments on our interdisciplinary paper.

Funding

The present study was funded by a grant from the Foundation for the Fusion of Science and Technology (Heisei27-10).

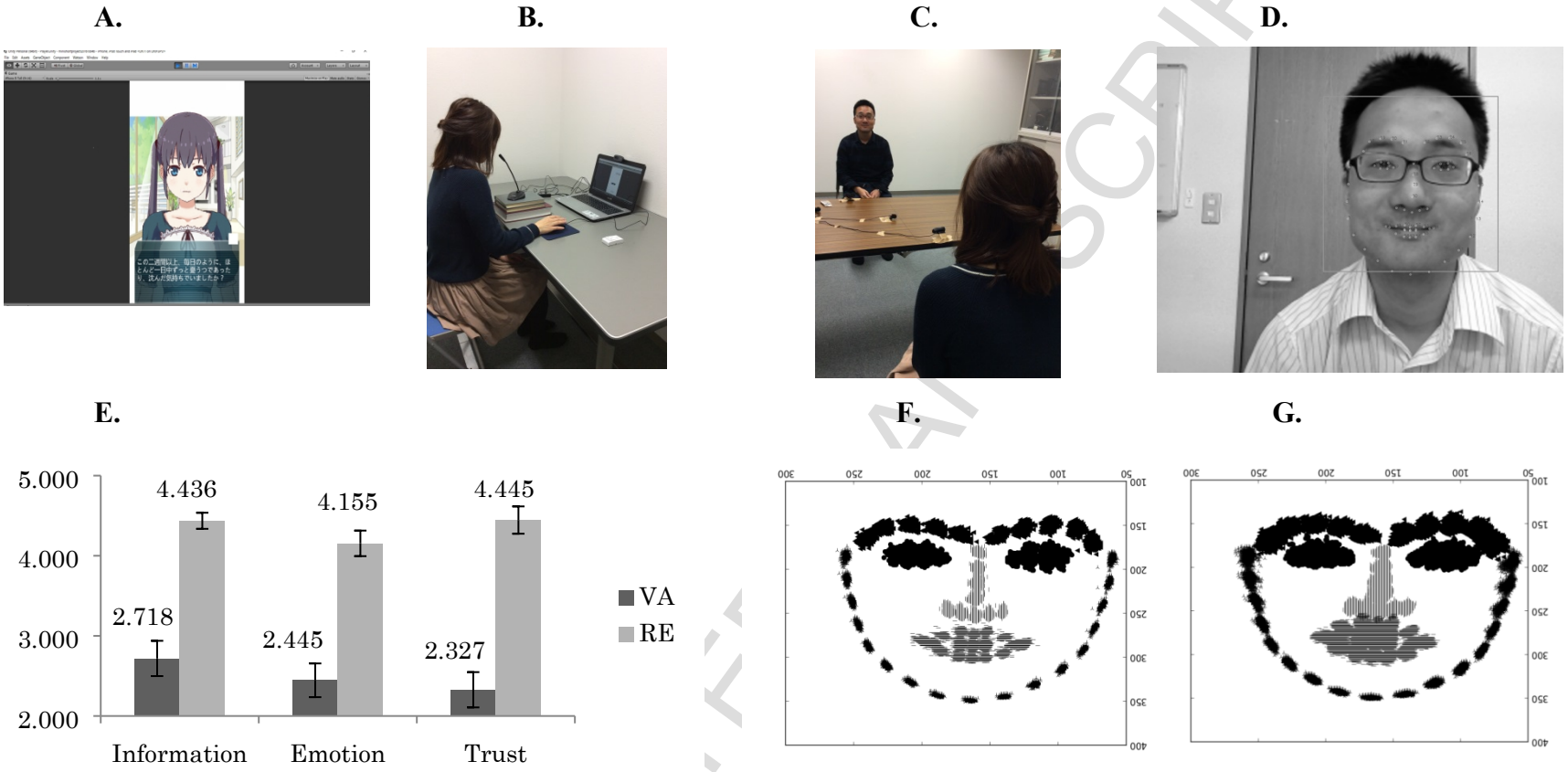

Figure 1. Comparison of the virtual agent (VA) and real expert (RE) conditions. VA appearance (A) and set-up (B), RE set-up (C), facial landmarks (D), gaps in perceived rapport with the VA versus that with the RE (E), facial movement during the VA (F), and RE set-up (G). In graph E, the margins of error concerning information, emotion, and trust during the VA scenario are 0.220, 0.210, and 0.220, respectively, and those during the RE scenario are 0.100, 0.160, and 0.170, respectively.

Highlights

- We clarified advantages of virtual agents and real experts in mental health.
- Virtual agent has advantages in participants' disclosure of sex-related symptoms.
- Real expert has advantages in participants' disclosure of mood-anxiety symptoms.
- Real expert has advantage of encouraging rapport and negative emotional expression.
- Virtual agent theories should distinguish real expert from real human.

Table 1.

Comparison of Conditions between Virtual Agent (VA) and Real Expert (RE)

		VA		RE	
Screening	Demographic Data	×		○	
	Educational and Work History	×		○	
	Past Mental History	×		○	
	Medical History	×		○	
	Current Symptoms	×		○	
	Current Social Functioning	×		○	
Training	Audio Adjustment	○		×	
Topics	Major Depressive Episode (MDE)	MINI A (P)		SCID A (C,L)	
	Categorization of MDE	MINI A (C) [a]		SCID A (C)	
	Manic/Hypomanic Episode	MINI D (L)		SCID A (C,L)	
	Categorization of Manic Episode	×		SCID A (C)	
	Dysthymic Disorder	MINI B (C)		SCID A (C)	
	Categorization of Mood Disorder	×		SCID D (C, L)	
	Alcohol Abuse	MINI J (C)		SCID E (C, L) [b]	
	Alcohol Dependence	MINI J (C)		SCID E (C, L) [b]	
	Panic Disorder	MINI E (C)		SCID F (C,L)	
	Agoraphobia	MINI F (C)		SCID F (C,L)	
	Specific Phobia	×		SCID F (C,L)	
	Social Phobia	MINI G(C)		SCID F (C,L)	
	OCD	MINI H(C)		SCID F (C,L)	
	PTSD	MINI I (C)		SCID F (C,L)	
	Generalized Anxiety Disorder	×		SCID F (C,L)	
	Anorexia	MINI M(C)		SCID H(C,L)	
	Bulimia	MINI N(C)		SCID H(C,L)	
Diagnosis Criteria	Not Directly Due to Physical Condition	×		○	
	Not Better Accounted for by Other Mental Disorders	×		○	
		M	*ME*	*M*	*ME*
	Duration of the Interview	20.76	1.40	38.23	4.49

Note: MINI: the Mini-International Neuropsychiatric Interview 5.0, A to N represents a module of the MINI. SCID: the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision Non-patient edition, A to H represents a module of the SCID. M: mean, ME: margin of error, OCD: Obsessive-Compulsive Disorder, C: current experience, P: past experience, L: lifetime experience. [a]: Melancholy only, [b]: Drug-related disorder was excluded.

Table 2.

Comparison of Areas of Facial Movement between Virtual Agent (VA) and Real Expert (RE) Conditions

		VA		RE		*F*	*Cohen's F*		Adjusted
	Area	*M*	*ME*	*M*	*ME*	*(1, 50)*		η_g^2	*p*
Eye	Right Eye (VM)	1.076	0.100	1.403	0.100	49.150	.991	.148	***
	Right Eye (HM)	0.893	0.066	1.022	0.070	11.082	.471	.058	*
	Left Eye (VM)	1.082	0.096	1.191	0.080	4.454	.298	.020	
	Left Eye (HM)	0.961	0.083	0.950	0.047	0.614	.111	.003	
Mouth	Inner Lip (VM)	1.238	0.100	1.380	0.089	6.813	.369	.037	
	Inner Lip (HM)	0.983	0.061	0.962	0.039	0.677	.116	.005	
	Outer Lip (VM)	1.205	0.096	1.278	0.077	2.002	.200	.012	
	Outer Lip (HM)	1.021	0.064	1.016	0.043	0.098	.044	.001	
Nose	Nasal Cavity (VM)	0.919	0.080	1.033	0.066	7.336	.383	.037	
	Nasal Cavity (HM)	0.685	0.049	0.660	0.033	2.046	.202	.009	
	Bridge of Nose (VM)	1.011	0.084	1.059	0.070	1.484	.172	.006	
	Bridge of Nose (HM)	0.723	0.071	0.703	0.052	1.454	.171	.004	
Brow	Right Eyebrow (VM)	1.383	0.170	1.370	0.095	0.074	.038	.000	
	Right Eyebrow (HM)	1.291	0.140	1.342	0.085	0.797	.126	.003	
	Left Eyebrow (VM)	1.417	0.160	1.366	0.110	1.587	.178	.004	
	Left Eyebrow (HM)	1.414	0.160	1.287	0.098	6.871	.371	.019	
Jaw	Jaw (VM)	1.264	0.062	1.414	0.065	16.334	.572	.085	**
	Jaw (HM)	0.847	0.038	0.837	0.030	0.690	.118	.003	

Note: $N = 54$, One participant's facial movements were disregarded because his fringe almost covered his brow and eyes, and our face detector could not determine his facial landmarks. Right and left indicate participant's right and left. VA: virtual agent, RE: real expert, M: mean, ME: margin of error, VM: vertical movement, HM: horizontal movement, F is score of F distribution (1, 50).The η_g^2 is generalized η squared (Olejnik & Algina, 2003). *Cohen's F* is $\sqrt{partial\ \eta^2/(1 - partial\ \eta^2)}$ (Cohen, 1988). The *p* value was adjusted through the Bonferroni method. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$

Table 3.

Comparisons of Self-disclosure for Each Mental Topic between the Virtual Agent (VA) and Real Expert (RE)

Topics	Disorders	TP RE(+)VA(+)	FP RE(-)VA(+)	FN RE(+)VA(-)	TN RE(-)VA(-)
Alcohol	Dependence (C) [a]	7	5	2	41
	Abuse (C) [a]	7	5	3	40
Mood	Major Depressive Disorder (C)	1	0	0	54
	Major Depressive Disorder (P)	0	0	13	42
	Depressive Symptoms with Melancholy (C)	0	1	0	54
	Dysthymic Disorder (C)	0	1	1	53
	Hypomanic Episode (L)	3	1	6	45
	Manic Episode (L)	0	0	0	55
Anxiety	Panic Disorder (L)	0	0	2	53
	Panic Disorder (C)	0	0	1	54
	Agoraphobia (L)	1	1	1	52
	Social Phobia (C)	0	0	5	50
Obsessive-Compulsive	Obsessive-compulsive Disorder (C)	0	0	1	54
Posttraumatic Stress	Posttraumatic Stress Disorder (C)	0	0	0	55
Eating	Anorexia Nervosa (C)	0	2	0	53
	Bulimia Nervosa (C)	0	2	0	53

Note: $N = 55$, (+): The disclosed mental symptoms are sufficient to diagnose, (-): The disclosed mental symptoms are insufficient to diagnose, TP: true positive, FP: false positive, FN: false negative, TN: true negative. C: current experience, P: past experience, L: lifetime experience, [a]: Current experience is defined within the last 12 months.