



# Automatic Thoughts and Facial Expressions in Cognitive Restructuring With Virtual Agents

Kazuhiro Shidara<sup>1\*</sup>, Hiroki Tanaka<sup>1</sup>, Hiroyoshi Adachi<sup>2</sup>, Daisuke Kanayama<sup>2</sup>, Yukako Sakagami<sup>2</sup>, Takashi Kudo<sup>2</sup> and Satoshi Nakamura<sup>1</sup>

<sup>1</sup> Nara Institute of Science and Technology, Nara, Japan, <sup>2</sup> Osaka University, Osaka, Japan

## OPEN ACCESS

### Edited by:

Gelareh Mohammadi,  
University of New South Wales,  
Australia

### Reviewed by:

Walter Gerbino,  
University of Trieste, Italy  
Everylyne Kimani,  
Northeastern University, United States

### \*Correspondence:

Kazuhiro Shidara  
shidara.kazuhiro.sc5@is.naist.jp

### Specialty section:

This article was submitted to  
Human-Media Interaction,  
a section of the journal  
Frontiers in Computer Science

**Received:** 21 August 2021

**Accepted:** 11 January 2022

**Published:** 02 February 2022

### Citation:

Shidara K, Tanaka H, Adachi H, Kanayama D, Sakagami Y, Kudo T and Nakamura S (2022) Automatic Thoughts and Facial Expressions in Cognitive Restructuring With Virtual Agents.  
*Front. Comput. Sci.* 4:762424.  
doi: 10.3389/fcomp.2022.762424

Cognitive restructuring is a well-established mental health technique for amending automatic thoughts, which are distorted and biased beliefs about a situation, into objective and balanced thoughts. Since virtual agents can be used anytime and anywhere, they are expected to perform cognitive restructuring without being influenced by medical infrastructure or patients' stigma toward mental illness. Unfortunately, since the quantitative analysis of human-agent interaction is still insufficient, the effect on the user's cognitive state remains unclear. We collected interaction data between virtual agents and users to observe the mood improvements associated with changes in automatic thoughts that occur in user cognition and addressed the following two points: (1) implementation of a virtual agent that helps a user identify and evaluate automatic thoughts; (2) identification of the relationship between a user's facial expressions and the extent of the mood improvement subjectively felt by users during the human-agent interaction. We focus on these points because cognitive restructuring by a human therapist starts by identifying automatic thoughts and seeking sufficient evidence to find balanced thoughts (evaluation of automatic thoughts). Therapists also use such non-verbal behaviors as facial expressions to detect changes in a user's mood, which is an important indicator for guidance. Based on the results of this analysis, we provide a technical guidance framework that fully automates the identification and evaluation of automatic thoughts to achieve a virtual agent that can interact with users by taking into account their verbal and non-verbal behaviors in face-to-face situations. This research supports the possibility of improving the effectiveness of mental health care in cognitive restructuring using virtual agents.

**Keywords:** cognitive restructuring, cognitive behavior therapy, virtual agent, automatic thought, facial expression

## 1. INTRODUCTION

Cognitive restructuring, an established therapeutic technique that reduces the effects of negative thoughts, is a cognitive behavior therapy (CBT) method (Beck and Beck, 2011). Some proposed virtual agents employ CBT methodology with mixed degrees of effectiveness (Laranjo et al., 2018; Montenegro et al., 2019). Several types of agent dialogues have also been proposed, such as text-based dialogues in the style of messaging apps (Fitzpatrick et al., 2017; Ly et al., 2017; Fulmer et al., 2018; Inkster et al., 2018; Sukanuma et al., 2018) and virtual agents represented by animated computer characters (Ring et al., 2016; Kimani et al., 2019). Virtual agents have many

advantages because they can provide face-to-face multimodal interactions like CBT with human therapists. In actual psychotherapy, the therapist understands the patient's condition not only through words but also through such non-verbal communication as facial expressions, speaking rhythms, and gestures (Koole and Tschacher, 2016). References to patients' facial expressions are also made in the CBT literature in clinical practices (Beck and Beck, 2011). If a virtual agent, like a psychiatrist, can recognize both verbal and non-verbal behaviors, it might approach the level of cognitive restructurings that humans do with each other. Therefore, dialogues with virtual agents can facilitate healthcare dialogues using non-verbal communication like facial expressions and voices, similar to human dialogues. According to a 2019 survey on the use of virtual agents in health contexts (Montenegro et al., 2019), some studies refer to medical datasets to study non-verbal behavior. For example, virtual agents have been proposed to provide social skills training based on the analysis of non-verbal behavior in autism spectrum disorders (Tanaka et al., 2017). Such cases remain rare, and to the best of our knowledge, no agent has been created based on analysis of the non-verbal features of cognitive restructuring.

In addition to virtual agents, CBT-based technical approaches have been proposed. Several studies have begun to report CBT adaptation to robot-to-human interactions. Dino et al. (2019) conducted a cohort study that provides CBT-based interactions to a small number of participants. Akiyoshi et al. (2021) applied CBT theory to promote self-disclosure in robot-to-human interactions. Human-to-human CBT within virtual reality has also been studied. Wallach et al. (2009) performed a randomized clinical trial of CBT on virtual reality. Lindner et al. (2019) presented a perspective on the usefulness and challenges of virtual reality in CBT.

Cognitive restructuring identifies the biased and useless ideas mistakenly used by patients to solve their current problems (Beck and Beck, 2011). In cognitive restructuring, a patient evaluates her automatic thoughts by evaluating their accuracy through a fact-based evaluation. Then the therapist asks some fact-finding questions and encourages the evaluation of automatic thoughts. Worksheets called thought records are often used in cognitive restructuring for personal use and allow therapists to reconstruct cognition. After the original thought record concept was proposed (Beck and Beck, 2011), Greenberger's updated version (Greenberger and Padesky, 2015) also became widely used. An important advance in Greenberger's thought records was to include Socratic questioning for evaluating a patient's expression of automatic thoughts. Questions are taken from Beck and Beck (2011) and can be widely used for general automatic thoughts.

CBT can treat such mental health conditions as anxiety and depression in members of the general population. In this paper, our system is intended for daily mental health care for members of the general population. Our final goal is to contribute to a mental health care system that provides cognitive behavior therapy to members of the general population before the onset of illness. In our data collection, our participants are graduate/undergraduate students who have not

been diagnosed with depressive disorders. We analyzed human-agent interactions and focused on the concept of automatic thoughts as used in cognitive restructuring. Automatic thoughts are interpretations that occur instinctively, depending on the situation. In the theory of cognitive restructuring, a situation and its reaction, such as moods, are not directly linked; automatic thoughts mediate them. In other words, the situation itself does not cause negative moods. Negative moods are caused by automatic thoughts that surface during specific situations. Cognitive restructuring improves a person's moods by confirming the bias of automatic thoughts concerning the situation and correcting them. In cognitive restructuring, the therapist first makes the patient aware of (identifies) automatic thoughts and then considers whether they are negative/biased or factual. If an automatic thought is negative/biased, effective cognitive restructuring can lead to valid/factual thoughts. Patients suffering from biased thoughts can improve their mood by eliciting new thoughts.

Since the reliable identification of automatic thoughts by patients/users is essential for evaluating them, we propose a system that helps users identify their automatic thoughts. When a virtual agent asks a user for automatic thoughts, it automatically determines whether their answers are automatic thoughts, and if it fails, this leads to successful identification.

The objective of this study is to observe the mood improvements associated with changes in automatic thoughts that occur in user cognition. We help users reliably identify automatic thoughts and investigate which questions in cognitive restructuring affect the user's moods during human-agent interaction. We hypothesize that an item called the evaluation of automatic thoughts affects the user's transformation and improves his negative moods.

This work makes the following two contributions:

1. Implemented a virtual agent that helps users identify and evaluate automatic thoughts.
2. Analyzed the relationship between user facial expressions and the extent of mood improvement subjectively felt by users during the human-agent interaction.

This paper is an extended version of our previously published works (Shidara et al., 2020, 2021). It synthesizes our prior works and includes some extensions and detailed explanations. Specifically, we extend the previous works by applying machine learning models to classify users' utterances about an automatic thought in Section 2.2. In addition, we added new participants in Section 3 to increase the experiment reliability on the relationship between the evaluation of automatic thoughts and mood improvement.

## 2. IDENTIFICATION OF AUTOMATIC THOUGHTS AND RELATIONSHIPS BETWEEN MOOD CHANGES AND FACIAL EXPRESSIONS

This section describes our data collection with a virtual agent system. Our data include the text, video, and speech of users. We

implemented an automatic thought classifier using the text data and analyzed the facial expressions of human-agent interactions.

The system's interface consists of five parts: a camera, a microphone, a keyboard, speakers, and a display. The camera is used only for face orientation and voice recording, and the microphone is used only for voice recording. The display projects the virtual agent from the chest up. The facial expressions and postures of the virtual agents are default settings. An interaction is composed of alternating questions from the virtual agent and the answers of one user. When the user presses the keyboard, the virtual agent's question is heard from the speaker. At this time, the lip-sync manager generates lip movements that match the virtual agent's voice. After listening to the virtual agent's question, the user answers it. After completing her answer, she presses a key on the keyboard for the next question.

After the data collection, we analyzed the mood strength, automatic thoughts, and changes in the users' facial expressions during their interactions with the virtual agent. We focused on reducing the variables involved in the interaction analysis as much as possible. In our data collection, the system asked fixed questions without commenting on the user responses.

We created a scenario based on cognitive restructuring (Beck and Beck, 2011). **Table 1** shows a system-side scenario created under the supervision of a psychiatrist (sixth author). We implemented this scenario in a virtual agent module with MMDAgent (Lee et al., 2013), which is a toolkit for building embedded conversational agents that allows voice dialogues. We used such default parameters as facial expressions, body posture, speaking speed, and voice pitch. The virtual agent outputs have spoken language, and the user inputs the natural language of the spoken language through a headset microphone.

## 2.1. Participants in Data Collection

We recruited 23 undergraduate/graduate students as users (10 females and 13 males). The research ethics committees of the Nara Institute of Science and Technology reviewed and approved this experiment (reference number: 2019-1-24-2). Written informed consent was obtained from all users before the experiment. The first author explained it to them and obtained their informed consent. We also confirmed that they had no severe mental illness based on the Kessler Psychological Distress Scale (K6) (Kessler et al., 2002). K6 is an inventory that does not distinguish between depression and anxiety disorders and investigates the tendency of mental illness. K6 scores range from 0 to 24, and the cutoff for our experiment inclusion was 13. All of the users' K6 scores were lower than the cutoff [mean ( $M$ ) = 7.18, standard deviation ( $SD$ ) = 4.22].

## 2.2. Analysis of Automatic Thoughts

In **Table 1**, Q4 identifies automatic thoughts. For example, the user's answer to Q4 is her automatic thought. After collecting the data, we extracted the answers to each user's Q4, and a psychiatrist who has clinical experience in cognitive restructuring labeled them based on whether they corresponded to an automatic thought. We identified the following three patterns in which the identification of automatic thoughts failed:

- Users confuse automatic thoughts with facts like "I can't decide what kind of job to look for."
- Users confuse their mood with such automatic thoughts as "I don't want to do anything."
- Users evaded answering questions with such responses as "Nothing" or "I don't understand."

If the user's answer to Q4 includes an automatic thought, its identification was deemed successful even if the user mentioned such collateral aspects as mood and situation. In the labeling, the psychiatrist determined that eight of 23 users failed to identify their automatic thoughts.

This study implemented a novel classification model. Based on the methodology of cognitive restructuring, we aim to build basic technology that enables virtual agents to shape automatic thoughts. Human therapists use a variety of questions to help patients who struggle to identify automatic thoughts. By identifying such thoughts through questions, the next step can be taken: their evaluation.

Evaluating an automatic thought means judging whether the automatic thought identified by the patient is based on negatively distorted cognition or factual validity. If an automatic thought is distorted, it can be evaluated and modified to suit a particular situation. Changing distorted thoughts to balanced thoughts generally improves a person's mood. On the other hand, cognitive restructuring does not improve a mood when an automatic thought is not negatively distorted. Therefore, after cognitive restructuring, action must be taken to solve the actual problem. Both therapists and patients can only determine if their automatic thoughts are distorted after undergoing cognitive restructuring.

Since the reliable identification of automatic thoughts by patients/users is essential for evaluating them, we propose a system that helps virtual agents identify the automatic thoughts of users. When a virtual agent asks users for automatic thoughts, this system automatically classifies whether their answers are automatic thoughts, and if the identification by users fails, this leads to successful identification. To realize this system, we built a classification model of sentences of automatic thoughts. By using the data (of the 23 participants) collected by the virtual agent to build a classification model, we achieved a highly practical classification performance.

We also added to the learning data an example sentence (**Table 2**) about automatic thoughts from a self-help book (Greenberger and Padesky, 2015). This list of example sentences is a worksheet with which general readers can acquire skills for identifying automatic thoughts. Thirty-three sentences are labeled as situations, moods, or thoughts. In CBT, situations and moods are different concepts from thoughts, although distinguishing among them is difficult unless one is familiar with them. We used thought as a successful identification and situation and mood as unsuccessful ones. **Table 3** shows the overview of the two datasets used to build the model.

The classifier algorithm was a support vector machine (linear kernel). The feature extraction method used in this paper's classification model has distributed representation by the Term Frequency-Inverse Document Frequency (TF-IDF) and Bidirectional Encoder Representations from Transformers

**TABLE 1** | Scenario of utterances spoken by system (translated from original Japanese into English).

No.	Item	Sentence (Virtual agent's questions)
Q1	Situation	<i>Hello, my name is Mai, and I'm a therapist. I'm here to help you learn how to face your worries. Has anything been bothering you lately, maybe something that's difficult to deal with or to face? Please describe anything that is painful or burdensome.</i>
Q2	Mood	<i>How did your mood change that time?</i>
Q3	Mood score (0–100%)	<i>How would you describe your mood's intensity from 0 to 100?</i>
Q4	Identification of an automatic thought	<i>I want to know why you felt like that in that situation. What thoughts came to you when you faced that event? I see...</i>
Q5	Proof	<i>Such thoughts that surface when we face certain events are called automatic thoughts. Since automatic thoughts are unconscious, they sometimes seem true. If your automatic thought is accurate, what do you think it is based on?</i>
Q6	Disproof	<i>So, do you have any other thoughts about that situation? For example, does anything contradict your automatic thoughts?</i>
Q7	Disputing(confirmation)	<i>Do you have any other thoughts? If so, can you share them?</i>
Q8	Socratic questioning* 1	<i>Let's examine the basis of the automatic thought that we have considered so far as well as how thoughts are balanced by contrary facts. Let me ask you some questions. What is the worst possible consequence of this situation?</i>
Q9	Socratic questioning 2	<i>Looking at the other side, what is the end result when you act as you want?</i>
Q10	Socratic questioning 3	<i>Please describe the most realistic scenario based on the predictions you just made.</i>
Q11	Balanced thought	<i>From that answer, let's create a new thought. How can you rethink this event?</i>
Q12	Caring words 1	<i>Good. Do you have anyone who will listen to you talk about this problem?</i>
Q13	Caring words 2	<i>I see, can you manage that?</i>
Q14	Mood score after change (0 to 100%)	<i>Well, our time is up for today. Thank you for your hard work. Please feel free to call again. Please feel free to talk to me. How has the intensity of your original mood changed? Express this on a scale from 0 to 100.</i>
[End]	Concluding a session	<i>If it's different from the beginning, that suggests that you internalized my suggestions well. That's it for today. Thank you for your hard work. Please feel free to call again.</i>

\*Questions that create awareness of inconsistency in thoughts.

(BERT) (Devlin et al., 2018). TF-IDF calculates importance by considering both the common and uncommon words in the dataset. When calculating TF-IDF, we used a morphological analyzer called MeCab<sup>1</sup> to divide the data. In vectorization, only the vocabulary in the training data is counted; unknown words are not counted.

BERT is a pre-trained language model for language comprehension. In this study, we tokenized the raw text and directly input the part-of-speech tagging for each token into the model. During tagging, special tokens are placed at the beginning and ending of sentences. A token named [CLS] is placed at the beginning of a sentence, and [SEP] is placed at its end. These tokens are placed in all input sentences. A sequence with a [CLS] token added to its beginning is input to BERT, and the output vector corresponding to the [CLS] token is used as the feature quantity. The BERT hidden vector has 768 dimensions.

**Table 4** shows the classification results. The best F1-score was 0.88 when we used sentences from Greenberger and Padesky (2015) as training data and TF-IDF as the features.

## 2.3. Relationship Between Mood Changes and Facial Expressions

### 2.3.1. Mood Scores and Mood Changes

We obtained the mood scores of participants by asking them twice about their negative mood intensities. The first score is Q3 and the second is Q14, both of which are shown in **Table 1**. We analyzed the changes in their moods and the corresponding changes in their facial expressions caused by the two ratings. We described the users' moods with such labels as anxious, depressed, sad, inferior, and fatigued. In this paper, we uniformly lumped all such feelings under the rubric of negative moods and only evaluated them by mood scores to focus on the changes themselves. Based on a related work (Persons and Burns, 1985), we calculated the mood changes using the following formula:

$$\text{Mood change} = \frac{(\text{mood score at beginning}) - (\text{mood score at end})}{(\text{mood score at beginning})} \quad (1)$$

In this research, we implemented a virtual agent that automatically interacted with users and collected and experimented with data. We used these interactions for the mood

<sup>1</sup><https://taku910.github.io/mecab/>

**TABLE 2** | Example sentences related to automatic thoughts (Greenberger and Padesky, 2015, p. 47–49).

	<b>Situation, mood, or thought (original labels in Greenberger and Padesky, 2015)</b>	<b>Successful or unsuccessful identification of an automatic thought (labels transformed to build a classification model)</b>
I'm not going to be able to do this	Thought	Successful
I'm going crazy	Thought	Successful
I'm no good	Thought	Successful
Something terrible is going to happen	Thought	Successful
Nothing ever goes right for me	Thought	Successful
I'll never get over this	Thought	Successful
I'm not going to be able to do this	Thought	Successful
I'm out of control	Thought	Successful
I'm a failure	Thought	Successful
She's so inconsiderate	Thought	Successful
I'm a loser	Thought	Successful
I feel like I'm having a panic attack	Thought	Successful
I've being taken advantage of	Thought	Successful
This isn't going to work out	Thought	Successful
I'm going to lose everything	Thought	Successful
At home	Situation	Unsuccessful
Talking to a friend on the phone	Situation	Unsuccessful
Driving in my car	Situation	Unsuccessful
At work	Situation	Unsuccessful
4:00 p.m.	Situation	Unsuccessful
Sitting in a restaurant	Situation	Unsuccessful
Talking to my mom	Situation	Unsuccessful
At my son's house	Situation	Unsuccessful
Lying in bed trying to sleep	Situation	Unsuccessful
Nervous	Mood	Unsuccessful
Sad	Mood	Unsuccessful
Annoyed	Mood	Unsuccessful
Frustrated	Mood	Unsuccessful
Discouraged	Mood	Unsuccessful
Depressed	Mood	Unsuccessful
Guilty	Mood	Unsuccessful
Ashamed	Mood	Unsuccessful
Panicky	Mood	Unsuccessful

scores to evaluate the dialogues. We did not use a standardized rating scale like K6 because our experiment was comprised of just one session. Such rating scales as K6 are unsuitable for assessing interactions in a single session because such evaluation scales assume that they will be re-measured at regular intervals. For example, K6 requires a one-month interval. Instead, a mood score is used in actual cognitive restructuring to measure mood improvement around a single session. Even though it is not a standardized measure, it is the most reasonable way to analyze one interaction session. Furthermore, a previous study (Persons and Burns, 1985) analyzed the cognitive restructuring of human therapists and patients and used mood scores as an evaluation index for a session. In our study, we calculated mood changes, which is the amount of change in the mood scores before and

after a session, by referring to Persons and Burns (1985) and using it as an evaluation index.

First, the users read a publically available explanation of CBT<sup>2</sup> that is designed for both clinical and general public uses. Then the first author explained how to use the system. Data were collected using a laptop PC (HP Probook) in a quiet room where the users sat alone. We recorded their facial expressions with its built-in camera. The completion time fluctuated based on the amount of speaking done by the user.

<sup>2</sup><https://www.cbtjp.net/downloads/skillup/pdf/>



**TABLE 3** | Datasets used for model building.

	Data collected in this study	Sentences from Greenberger and Padesky (2015)
Number of sentences	23	33
Annotator	6th author of this paper (TK)	Authors of Greenberger and Padesky (2015)
Sentence style	Speech transcription	Written language
Usage in model building	Training and test	Training Mood/feeling/thought (We replaced thought with the successful identification of automatic thought, and situation and feeling with the unsuccessful identification of an automatic thought.)
Label	Successful/unsuccessful identification of an automatic thought	

**TABLE 4** | Classification results of thought: D: data collected in this study, G: sentences from Greenberger and Padesky (2015), Bold: best score.

Training data	Test data	Feature	Accuracy	F1-score
D + G	D	TF-IDF	0.70	0.80
		BERT	0.65	0.78
		TF-IDF + BERT	0.61	0.76
D	D	TF-IDF	0.61	0.76
		BERT	0.65	0.79
		TF-IDF + BERT	0.65	0.79
G	D	TF-IDF	<b>0.83</b>	<b>0.88</b>
		BERT	0.70	0.81
		TF-IDF + BERT	0.70	0.81

**TABLE 5** | Action units (AUs) for analysis.

AU	Description	AU	Description
AU1	Inner brow raiser	AU14	Dimpler
AU2	Outer brow raiser	AU15	Lip corner depressor
AU4	Brow lowerer	AU17	Chin raiser
AU5	Upper lid raiser	AU20	Lip stretcher
AU6	Cheek raiser	AU23	Lip tightener
AU7	Lid tightener	AU25	Lips part
AU9	Nose wrinkler	AU26	Jaw drop
AU10	Upper lip raiser	AU45	Eyes closed (blinking)
AU12	Lip corner puller		

### 2.3.2. Extracting Facial Action Units

Based on the Facial Action Coding System (FACS) (Ekman and Friesen, 1978), a quantitative method for describing facial movements, our analysis used action units (AUs), which are partial units of facial expressions. AUs were automatically extracted with OpenFace (Baltrušaitis et al., 2016), a facial expression analysis tool. We analyzed the 17 AUs listed in **Table 5**. The extracted AUs were represented as intensity information of continuous values from 0 to 5.

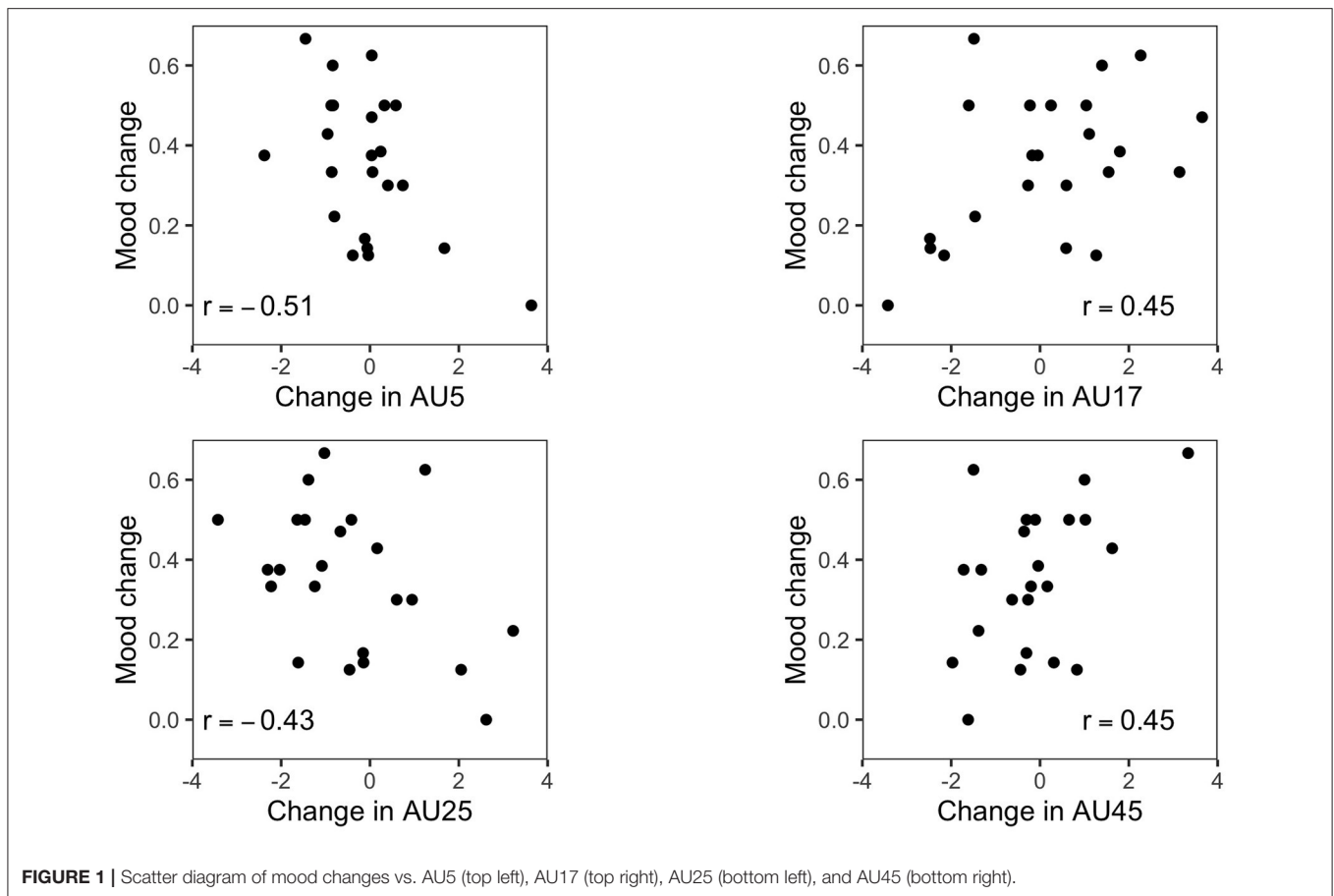
We extracted 17 types of AUs for each facial part from the video. Data processing for each AU was done independently. The data processing process for each AU is shown below:

- The AUs were extracted from the videos.
- OpenFace calculated the reliability of the successful face recognitions. Frames with less than 70% reliability were removed.
- The virtual agent and user turns were separated, and the virtual agent turns were ignored.
- The average AUs were calculated in the frame for each extracted user turn. Since there are 14 user turns (**Table 1**), we calculated the mean value of 14 types per user.

- The AU for each turn within the users was standardized with the mean value of the 14 AU types. This normalization calculated the relative AU strength for each turn within the users.
- The facial expression differences were calculated through the interactions by the following equation. The AU at the beginning is the period from the user's response turn to Q3 in the fixed scenario of **Table 1** (from the beginning of the user's utterance to its end). The AU at the end is the period of the user's response to Q14 in **Table 1** (from the start of the user's utterance to its end):

$$\text{Change of AU} = (\text{AU at the beginning}) - (\text{AU at the end}). \quad (2)$$

We analyzed the correlation between the mood and AU changes to clarify the correlation between the relative fluctuations in the mood changes and in the AUs. The AUs at the beginning and the end in Equation (2) are pre-relativized values within the individual. Therefore, the AU change in Equation (2) is a relative scale.



**FIGURE 1** | Scatter diagram of mood changes vs. AU5 (top left), AU17 (top right), AU25 (bottom left), and AU45 (bottom right).

### 2.3.3. Pearson's Correlation Analysis

The following changes in AUs were significantly correlated with mood changes at  $\alpha = 0.05$ : AU5 ( $r = -0.51$ ,  $t = -2.68$ ,  $p = 0.01$ ), AU17 ( $r = 0.45$ ,  $t = 2.32$ ,  $p = 0.03$ ), AU25 ( $r = -0.43$ ,  $t = -2.24$ ,  $p = 0.04$ ), and AU45 ( $r = 0.45$ ,  $t = 2.32$ ,  $p = 0.03$ ). **Figure 1** shows scatter diagrams between these changes in AUs and moods. AU5 and AU45 are squinting movements in the face's upper half; AU17 and AU25 are movements in its lower half, showing a raised chin. Positive correlations indicate that the greater the mood change, the lower its AU expression. A negative correlation indicates that the larger the mood change is, the greater the expression of that AU before and after.

## 3. EVALUATING QUESTIONS FROM THE VIRTUAL AGENT

This section describes our experiment using a virtual agent system. We investigated the effect of the evaluation of automatic thoughts in human-agent interaction and analyzed the relationship between the number of questions that users felt useful and their mood changes.

In the experiments in this section, we improved the system configuration and the scenario of the virtual agent. We employed a 3D animated virtual agent named Shibata (**Figure 2**) that uses

a virtual agent platform, Greta (Niewiadomski et al., 2009), which was modified in-house for Japanese text-to-speech, lip-synching, and Japanese-style animation (Tanaka et al., 2021). This virtual agent is autonomous and outputs questions through a voice. Then the user inputs natural language through a headset microphone.

**Figure 3** shows the configuration of our virtual agent. Its interface consists of five parts: a camera, a microphone, a speaker, and a display. The camera is used only for recording facial expressions. The microphones are used for voice recording, turn changes, and question selection. An interaction alternates one question from the virtual agent and one user answer. After the virtual agent asks a question, a speech recognition API is activated while the user answers. When the user stops talking for a certain period, the speech recognition API automatically recognizes the end. After obtaining the user's answer, the scenario selects a question from the scenario based on the question management rule. The question text is played back from the speaker as the voice of the virtual agent through text-to-speech. After listening to the virtual agent's question, the user answers it. The display visualizes the virtual agent by showing area above its chest. The facial expressions and postures of the virtual agent are its default settings. It performs no facial expressions or gestures for communication. The lip model and lip blender generate lip movements to match the voice of the virtual agent.

### 3.1. Dialogue Scenario

#### 3.1.1. Questioning to Identify Automatic Thoughts

Table 6 shows our experimental dialogue scenario. Patients sometimes fail to identify automatic thoughts in cognitive restructuring because this concept is not conventionally recognized (Beck and Beck, 2011). We began our investigation of this guiding function by investigating the percentage of user failures to identify automatic thoughts in cognitive restructuring with virtual agents. We automated the guide to identify a thought by the following procedure. Dialogue control with a classifier is done in Q6 of Table 6.

Experimental scenarios (Table 6) are created and generated after the data collection scenario (Table 6). The most crucial difference is that the experimental scenarios were improved for comparison with and without “questions for evaluating automatic thoughts.” In Table 6, words of concern (e.g., Q12 and

Q13) were included between Q5 (identification of an automatic thought) and Q14 (mood score after change). On the other hand, Table 6 includes only questions for evaluating an automatic thought between Q5 (identification of an automatic thought) and Q14 (mood score after change). The questions for evaluating an automatic thought in Table 6 were all taken from the same source (Beck and Beck, 2011).

1. The classifier determines whether the user identified a thought.
2. **Successful:** Move to the next item.
3. **Unsuccessful:** Ask a question to guide identification of the thought and ask the user to answer again.
4. **Unsuccessful:** Make a judgment with the automatic thought classifier. If it is unsuccessful again, provide another hint.

If the user fails to identify the automatic thought, the virtual agent provides a hint and asks the user to answer again: *So, what did you think about yourself in that situation?* Six hints are available, and up to six unsuccessful attempts to identify the automatic thought can be handled. The six hints are given to all users in the same order. If a user fails seven times, he must proceed to the next item. The hints are taken from Beck and Beck (2011).

#### 3.1.2. Questioning to Evaluate and Modify Automatic Thoughts

We evaluated automatic thoughts using the nine questions (Q6–Q14) shown in Table 6. We prepared two dialogue scenarios: one that asked all the questions and another that omitted questions Q7–Q13. Our experiment hypothesizes that evaluating automatic thoughts improves a user’s mood. We prepared these seven questions to guide the evaluation of automatic thoughts and asked them in a round-robin style, assuming that one of them would be effective for the user. To clarify the hypothesis, we compared two scenarios (Figure 4): Group A, questions



FIGURE 2 | Shibata, our virtual dialogue agent based on CBT.

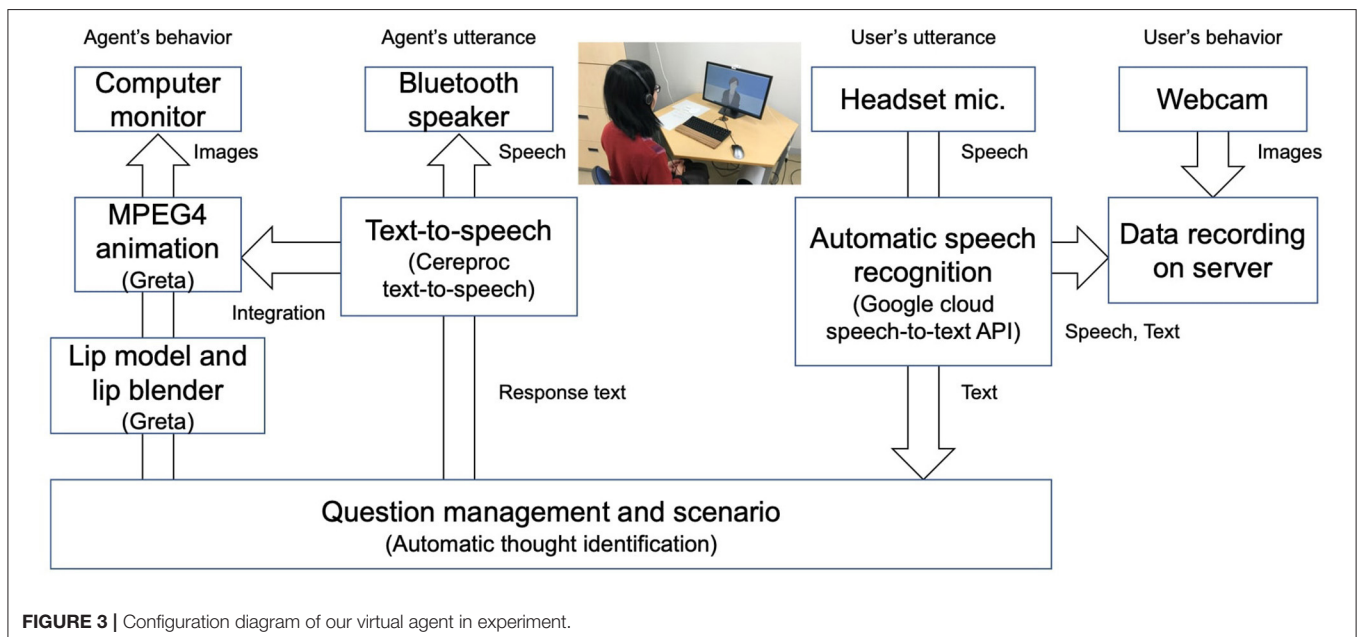


FIGURE 3 | Configuration diagram of our virtual agent in experiment.



**TABLE 6** | Scenario of utterances spoken by virtual agent and fictional user (we translated the original Japanese into English).

No.	Item	Sentence (V denotes a virtual agent, and P denotes a fictional user.)
Q1	Situation	V: Hello, my name is Shibata, and I'm a therapist. I'm here to help you with any worries you might have. Has anything been bothering you lately that's difficult to deal with or to face? Please tell me about anything that is painful or burdensome. P: Although I sent her an email, I haven't gotten a reply.
Q2	Mood	V: How did your mood change at that time? P: I was so sad.
Q3	Mood score (0–100%)	V: How would you describe the intensity of your mood from 0 to 100? P: 75.
Q4	Question to identify automatic thoughts	V: What thoughts came to you at that time? For example, "My coworkers hate me," "My father doesn't understand me," or "I might be fired." P: He hates me.
Q5	Check whether there are other automatic thoughts	V: I see. Is there anything else? If you've said everything you can think of, please say, "that's all." P: That's all.
Q6	Question to evaluate and modify a thought: Set up an automatic thought for the agenda	V: What is the strongest thought that comes to you? P: I'm hated.
Q7	Question to evaluate and modify thought 1	V: If your automatic thought is true, what evidence do you have to support this? P: I didn't get a reply after one night.
Q8	Question to evaluate and modify thought 2	V: If your automatic thought is not true, what evidence do you have to support this? P: My friends haven't told me they don't like me to my face yet.
Q9	Question to evaluate and modify thought 3	V: Besides your automatic thoughts, are there any other possible explanations? P: The email wasn't really urgent, and no particular damage has been done.
Q10	Question to evaluate and modify thought 4	V: What's the effect if you believe your automatic thought? P: I'll avoid getting involved with her.
Q11	Question to evaluate and modify thought 5	V: I see. So, what's the most realistic outcome? P: She may not be interested in me, but that doesn't mean she doesn't like me.
Q12	Question to evaluate and modify thought 6	V: I see. If your friend were in this situation and had this thought, what would you tell him or her? P: Why don't you wait for a response before taking it too seriously?
Q13	Question to evaluate and modify thought 7	V: What action would you take about your automatic thoughts? P: I'm going to reply without concern and wait.
Q14	Question to evaluate and modify thought: Balanced thought	V: In addition to the automatic thoughts you mentioned, maybe you can find another idea. If you think of a new thought, try connecting it to your automatic thoughts with "but..." P: Since he hasn't responded yet means that I don't think he likes me. But I'm not sure yet, so I'm going to take it easy and talk about the e-mail directly when I see him again.
Q15	Mood score after change (0–100%)	V: How has the intensity of your original mood changed? Express it on a scale from 0 to 100. P: 40.
[End]	Concluding a session	V: If it's different from the beginning, that suggests that you internalized my suggestions well. Our time is up for today. Thank you for your hard work. Please feel free to call again.

that evaluate thoughts; Group B, no questions that evaluate automatic thoughts.

## 3.2. Experimental Evaluation

### 3.2.1. Participants

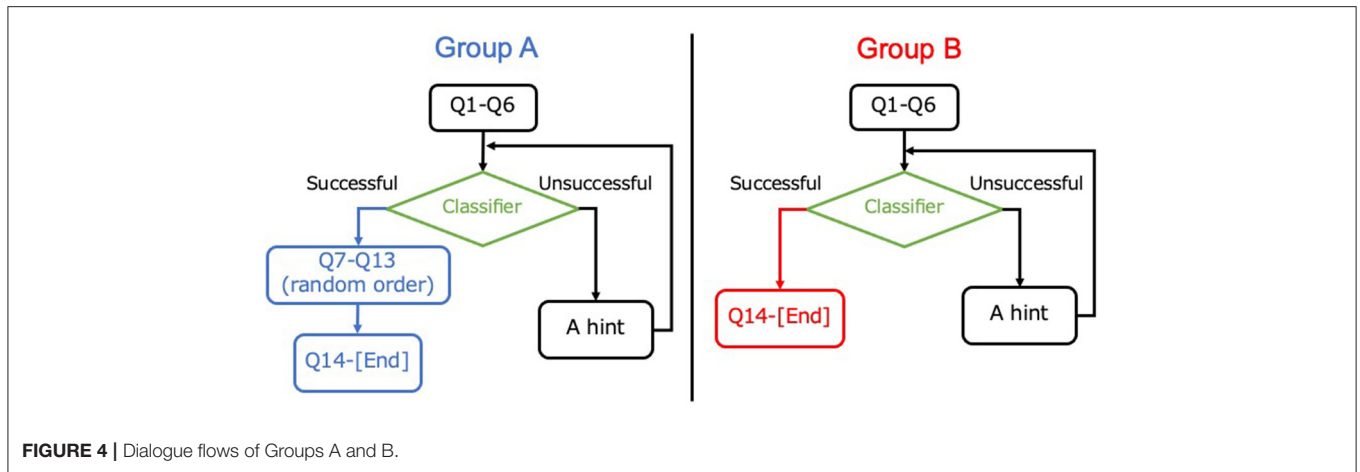
We recruited 32 graduate students as users: 19 in Group A and 13 in Group B. The research ethics committee of the Nara Institute of Science and Technology reviewed and approved this experiment (reference number: 2019-I-24-2). Written informed consent was obtained from all the users in it. We also confirmed with K6 tests that they suffered from no severe mental illness

tendencies. The K6 score for Group A was  $M = 4.74$ ,  $SD = 4.48$ , and for Group B it was  $M = 4.46$ ,  $SD = 2.57$ . There was no significant difference in the K6 scores between the two groups.

### 3.2.2. Procedure

The same procedure was applied to both groups. First, the users read a publically available leaflet that explained CBT<sup>3</sup> and how it is designed for both clinical and general applications. Next

<sup>3</sup><https://www.cbtjp.net/downloads/skillup/pdf/>



**TABLE 7** | Differences of virtual agent interface and scenario settings in data collection and experiment (Groups A and B).

		Data collection	Group A	Group B
Interface	Platform	MMDAgent	Greta	Greta
	Turn change	User manual	Automatic	Automatic
	Agent's input	Speech	Speech	Speech
	Agent's output	Speech	Speech	Speech
	Agent's visual	Japanese	Japanese	Japanese
	Agent's facial action	Default setting	Default setting	Default setting
	Agent's gesture	Default setting	Default setting	Default setting
	Language	Japanese	Japanese	Japanese
	Scenario	Guidance of identification of automatic thought	Not included	Included
Guidance of identification of automatic thought		Included	Included (modified from data collection)	Not included
Participant	Number of participants	23	19	13
	Mean K6 scores	7.18	4.74	4.46
	Standard deviation of K6 scores	4.22	4.48	2.57
	Experiment setting	Laboratory	Laboratory	Laboratory
	Recruitment target	Graduate or undergraduate students	Graduate students	Graduate students

we explained how to use the virtual agent. Our experiment was carried out on a desktop personal computer.

Unfortunately, Q2 (mood) in **Table 6** was skipped in the interaction of some participants. This mistake affected ten people in Group A and ten people in Group B. Since the scenario was not disclosed to the participants, they were unaware that Q2 was skipped. However, not all of the participants who experienced this situation were confused by its omission. For Q3, all participants clearly stated their own negative mood

types. Therefore, we also used the data of the participants who experienced this trouble along with the data of those who did not.

We performed an unpaired *t*-test on the mood changes of the two groups. The change is the relative fluctuation of the mood score. We chose the unpaired *t*-test over a two-way ANOVA (analysis of variance) because this experiment compares mood change values rather than mood score differences. **Table 7** shows the differences between the data collection and the interfaces and the scenarios of Groups A and B in this session's experiments.

## 4. RESULTS

### 4.1. Mood Changes Through Cognitive Restructuring

Figure 5 shows the improvement of the negative moods in both groups. The questioning with which the automatic thoughts were evaluated significantly affected the increases in the negative mood scores ( $p = 0.036$ , Hedge's  $g = 0.769$ ). In Group A, the following are the results: the mood score at the start was  $M = 55.0$ ,  $SD = 23.3$ ; the mood score at the end was  $M = 34.1$ ,  $SD = 18.6$ ; the mood change was  $M = 0.38$ ,  $SD = 0.22$ . In Group B, the following are the results: the mood score at the start was  $M = 42.8$ ,  $SD = 26.5$ ; the mood score at the end was  $M = 32.2$ ,  $SD = 18.0$ ; the mood change was  $M = 0.13$ ,  $SD = 0.32$ . We found no significant difference in the mood intensity between the two groups at the beginning and the end of the study.

### 4.2. Questions Deemed Helpful by Participants

Group A evaluated questions Q6–Q14 in Table 6 as helpful for discovering new thoughts. At the end of each user's dialogue, the users were given a questionnaire that contained the virtual agent's questions and rated each question as either helpful or unhelpful.

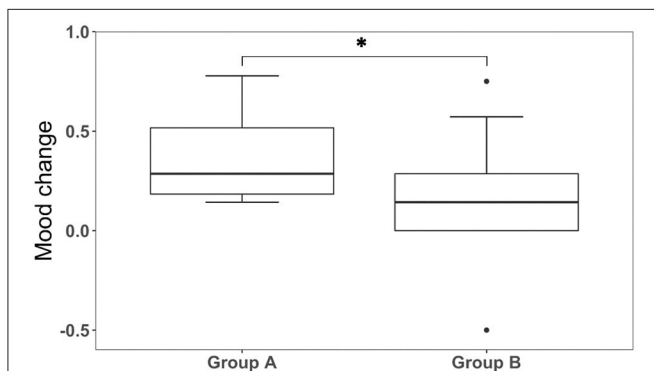


FIGURE 5 | Boxplot of changes in mood scores between Groups A and B. The symbol \* means the significance level  $\alpha = 0.05$  in the unpaired  $t$ -test.

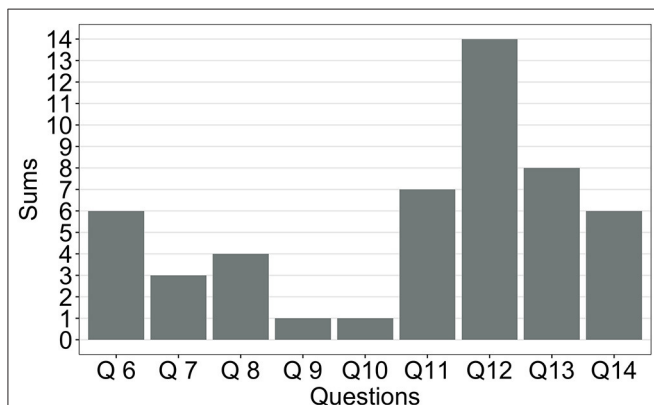


FIGURE 6 | Distribution of helpful questions ( $n = 19$ , multiple answers possible).

Figure 6 shows the distribution of the helpful questions. Q12 and Q13 were rated the most helpful by users; Q9 and Q10 were rated the least useful. One reason for this result is the difference in the intentions of the questions. Q9 and Q10 asked the users to dig deeper into their automatic thoughts. On the other hand, Q12 and Q13 examined automatic thoughts from a new perspective. Perhaps questions that change the user's perspective contribute more to modifying automatic thoughts.

### 4.3. Correlation Between Mood Scores and Number of Helpful Questions

We investigated the correlation between the mood scores and the number of helpful questions for Group A in the evaluation of automatic thoughts. Immediately after each interaction, the users filled out a questionnaire and answered whether they felt each question was helpful.

We calculated the correlation between the number of helpful questions for each user and the mood changes. We performed a Spearman's rank correlation coefficient test for these analyses. The results showed a strong correlation between the number of helpful questions and mood changes ( $\rho = 0.63$ ,  $p = 0.0035$ ; Figure 7).

## 5. DISCUSSION

The dialogues of the virtual agent using Table 1 significantly improved the subjects' negative moods. Our work is the first to analyze how facial movements, which are related to mood changes, are affected by interaction with a virtual agent. In addition, we identified the AUs that are probably influenced by mood changes. Our finding is expected to contribute to research on virtual agents that recognize facial movements and the basic research of human facial expression analysis. The correlation

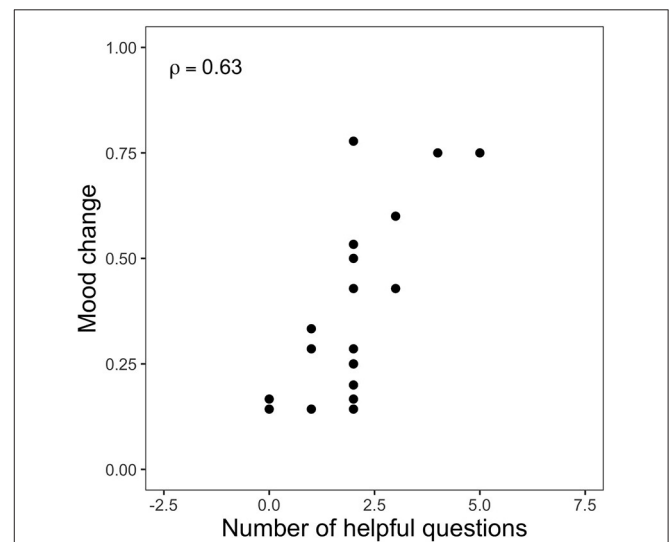


FIGURE 7 | Scatter plot between number of helpful questions and mood changes.

of facial action units and mood changes implies that cognitive restructuring care with a virtual agent can change users' moods, and such changes appear as facial expressions in proportion to mood improvements.

However, since it remains unclear whether virtual agents are better suited to promoting the expression of moods, analyzing facial expressions among different dialogue styles is another step in our future work. Since interpreting the reasons for these movements was complicated, we also reviewed the recorded videos. Consequently, those who displayed relatively large mood changes appeared to think more deeply when answering the mood score at Q14 than at Q3 in **Table 1**. Therefore, we assumed that a contemplative attitude was reflected in raised eyelids and closed mouths. Although we analyzed individual AUs, there was insufficient information to conclude that they were due to mood changes. Therefore, to recognize mood changes more reliably, both facial expressions and such multimodal behavioral indicators as voice and gestures must be analyzed. We have not yet fully verified that the correlation between mood and AU changes is due to factors other than mood.

This study also showed that using questions to evaluate automatic thoughts improved users' moods. Previous studies on CBT interactions (e.g., Fitzpatrick et al., 2017) generally investigated users' moods and depressive tendencies by comparing users with and without virtual agents. However, the factors that influence the effectiveness of such dialogues have not been sufficiently examined. We found that questioning that evaluates automatic thoughts is an important factor in improving negative moods. Furthermore, the more questions that were helpful to the users, the more their moods improved. Asking questions seems to provide new information to users. If gathering information worked well, users probably improved their moods. Therefore, our virtual agent suggests the effectiveness of leveraging the ability of guided modifications of automatic thoughts.

Finally, after experimenting using a virtual agent with a function that expresses automatic thoughts, we obtained an effect

that improved moods by evaluation questions. After analyzing the questions to evaluate automatic thoughts, the questions, which were intended to find different viewpoints, tended to be evaluated as more helpful for changing thoughts. The more questions that were helpful, the better the users' moods became. Our results suggest that a virtual agent's effectiveness can be improved by predicting helpful questions.

## DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because, the dataset created in this study is private. Requests to access the datasets should be directed to Kazuhiro Shidara, shidara.kazuhiro.sc5@is.naist.jp.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Nara Institute of Science and Technology reviewed and approved this experiment (reference number: 2019-I-24-2). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

KS, HT, TK, and SN conceived the entire experiment's design. KS, HT, and SN performed the experiments, the data analysis, and the data collection. KS, HT, and SN analyzed our data. All authors discussed the results. HA, DK, YS, and TK provided information on the field of CBT. KS wrote the manuscript. SN is the project's principal investigator and directs all the research. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by JST CREST Grant Number JPMJCR19A5, Japan.

## REFERENCES

- Akiyoshi, T., Nakanishi, J., Ishiguro, H., Sumioka, H., and Shiomi, M. (2021). A robot that encourages self-disclosure to reduce anger mood. *IEEE Robot. Autom. Lett.* 6, 7925–7932. doi: 10.1109/LRA.2021.3102326
- Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2016). "Openface: an open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Lake Placid, NY), 1–10. doi: 10.1109/WACV.2016.7477553
- Beck, J. S., and Beck, A. (2011). *Cognitive Behavior Therapy*. New York, NY: Guilford Publication.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dino, F., Zandie, R., Abdullahi, H., Schoeder, S., and Mahoor, M. H. (2019). "Delivering cognitive behavioral therapy using a conversational social robot," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau), 2089–2095. doi: 10.1109/IROS40897.2019.8968576
- Ekman, P., and Friesen, W. V. (1978). Facial Action Coding System (FACS). *PsycTests*. doi: 10.1037/t27734-000
- Fitzpatrick, K. K., Darcy, A., and Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR Ment. Health* 4:e19. doi: 10.2196/mental.7785
- Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., and Rauws, M. (2018). Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment. Health* 5:e64. doi: 10.2196/mental.9782
- Greenberger, D., and Padesky, C. A. (2015). *Mind Over Mood: Change How You Feel by Changing the Way You Think (Japanese edition)*. Guilford Publications.
- Inkster, B., Sarda, S., and Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth uHealth* 6:e12106. doi: 10.2196/12106
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L., et al. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol. Med.* 32, 959–976. doi: 10.1017/S0033291702006074
- Kimani, E., Bickmore, T., Trinh, H., and Pedrelli, P. (2019). "You'll be great: virtual agent-based cognitive restructuring to reduce public speaking anxiety," in *2019*

- 8th International Conference on Affective Computing and Intelligent Interaction (ACII) (Cambridge), 641–647. doi: 10.1109/ACII.2019.8925438
- Koole, S. L., and Tschacher, W. (2016). Synchrony in psychotherapy: a review and an integrative framework for the therapeutic alliance. *Front. Psychol.* 7:862. doi: 10.3389/fpsyg.2016.00862
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., et al. (2018). Conversational agents in healthcare: a systematic review. *J. Am. Med. Inform. Assoc.* 25, 1248–1258. doi: 10.1093/jamia/ocy072
- Lee, A., Oura, K., and Tokuda, K. (2013). “Mmdagent-a fully open-source toolkit for voice interaction systems,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (Vancouver, BC), 8382–8385. doi: 10.1109/ICASSP.2013.6639300
- Lindner, P., Hamilton, W., Miloff, A., and Carlbring, P. (2019). How to treat depression with low-intensity virtual reality interventions: perspectives on translating cognitive behavioral techniques into the virtual reality modality and how to make anti-depressive use of virtual reality-unique experiences. *Front. Psychiatry* 10:792. doi: 10.3389/fpsyg.2019.00792
- Ly, K. H., Ly, A.-M., and Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Intern. Intervent.* 10, 39–46. doi: 10.1016/j.invent.2017.10.002
- Montenegro, J. L. Z., da Costa, C. A., and da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Syst. Appl.* 129, 56–67. doi: 10.1016/j.eswa.2019.03.054
- Niewiadomski, R., Bevacqua, E., Mancini, M., and Pelachaud, C. (2009). “Greta: an interactive expressive ECA system,” in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Vol. 2* (Budapest: Citeseer), 1399–1400.
- Persons, J. B., and Burns, D. D. (1985). Mechanisms of action of cognitive therapy: the relative contributions of technical and interpersonal interventions. *Cogn. Ther. Res.* 9, 539–551. doi: 10.1007/BF01173007
- Ring, L., Bickmore, T., and Pedrelli, P. (2016). “An affectively aware virtual therapist for depression counseling,” in *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI) Workshop on Computing and Mental Health* (San Jose, CA).
- Shidara, K., Tanaka, H., Adachi, H., Kanayama, D., Sakagami, Y., Kudo, T., et al. (2020). “Analysis of mood changes and facial expressions during cognitive behavior therapy through a virtual agent,” in *Companion Publication of the 2020 International Conference on Multimodal Interaction* (Virtual event), 477–481. doi: 10.1145/3395035.3425223
- Shidara, K., Tanaka, H., Adachi, H., Kanayama, D., Sakagami, Y., Kudo, T., et al. (2021). “Relationship between mood improvement and questioning to evaluate automatic thoughts in cognitive restructuring with a virtual agent,” in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)* (Virtual event). doi: 10.1109/ACIIW52867.2021.9666312
- Suganuma, S., Sakamoto, D., and Shimoyama, H. (2018). An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR Ment. Health* 5:e10454. doi: 10.2196/10454
- Tanaka, H., Iwasaka, H., Matsuda, Y., Okazaki, K., and Nakamura, S. (2021). Analyzing self-efficacy and summary feedback in automated social skills training. *IEEE Open J. Eng. Med. Biol.* 2, 65–70. doi: 10.1109/OJEMB.2021.3075567
- Tanaka, H., Negoro, H., Iwasaka, H., and Nakamura, S. (2017). Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLoS ONE* 12:e182151. doi: 10.1371/journal.pone.0182151
- Wallach, H. S., Safir, M. P., and Bar-Zvi, M. (2009). Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. *Behav. Modif.* 33, 314–338. doi: 10.1177/0145445509331926
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Shidara, Tanaka, Adachi, Kanayama, Sakagami, Kudo and Nakamura. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.