

# Exploring User Experience and Performance of A Tedious Task Through Human-Agent Relationship

Chao Zhou<sup>1,+</sup>, Yulong Bian<sup>1,+</sup>, Shu Zhang<sup>1</sup>, Ziyang Zhang<sup>3</sup>, Yaoyuan Wang<sup>3</sup>, and Yong-Jin Liu<sup>1,2,\*</sup>

<sup>1</sup>Tsinghua University, the Department of Computer Science and Technology, Beijing, 100084, China

<sup>2</sup>Tsinghua University, BNRist, MOE-Key Laboratory of Pervasive Computing, Beijing, 100084, China

<sup>3</sup>Huawei Technologies Co Ltd, Advanced Computing and Storage Laboratory, Shenzhen, 518000, China

\*Corresponding author. liuyongjin@tsinghua.edu.cn

+These authors contributed equally to this work

## ABSTRACT

Positive human-agent relationships can effectively improve human experience and performance in human-machine systems or environments. The characteristics of agents that enhance this relationship have garnered attention in human-agent or human-robot interactions. In this study, based on the rule of the persona effect, we study the effect of an agent's social cues on human-agent relationships and human performance. We constructed a tedious task in an immersive virtual environment, designing virtual partners with varying levels of human likeness and responsiveness. Human likeness encompassed appearance, sound, and behavior, while responsiveness referred to the way agents responded to humans. Based on the constructed environment, we present two studies to explore the effects of an agent's human likeness and responsiveness to agents on participants' performance and perception of human-agent relationships during the task. The results indicate that when participants work with an agent, its responsiveness attracts attention and induces positive feelings. Agents with responsiveness and appropriate social response strategies have a significant positive effect on human-agent relationships. These results shed some light on how to design virtual agents to improve user experience and performance in human-agent interactions.

## Introduction

An intelligent agent is a software entity that can conduct operations continuously and autonomously after sensing a specific environment<sup>1</sup>. As AI technology rapidly progresses and becomes more widespread, intelligent agents have started appearing more frequently in our daily lives. The co-existence of people and these agents significantly promotes the study of human-agent interaction<sup>2</sup>. Agents are often given roles (for example, the roles of the teacher and partner<sup>3</sup>) based on which agents interact with people and assist people through virtual activities<sup>4</sup>.

User experience and performance are important factors in evaluating human-agent interaction<sup>2,5,6</sup>. Thus far, most existing human-agent interaction studies have focused on task performance efficiency without considering the human-agent relationship<sup>7,8</sup>. Despite achieving high-performance efficiency, forcing people to adhere to the repetitive operations of the machine can cause fatigue and a poor user experience<sup>9-11</sup>. In this study, we investigate user experience and performance in human-agent interaction using a tedious task from the perspective of positive human-agent relationships.

To the best of the authors' knowledge, "tedious tasks" have not yet been clearly defined in the literature. Based on the description of tedious activity in previous literature<sup>9,10</sup>, we define a "tedious task" as an activity or piece of work that the processes are unnecessarily monotonous, repetitive, and time-consuming. Performing a tedious task in human-agent interaction easily leads to a negative user experience<sup>9,10</sup>, and even reduces human performance and motivation to participate in the task. Therefore, it is desirable to explore effective factors to improve the human experience in tedious tasks by establishing a positive human-agent relationship.

Pedagogical agents have been widely studied in the literature as special agents. The presence of an animated pedagogical agent can lead to positive effects in learning activities, which has been well known as *Persona Effect*<sup>12,13</sup>. Furthermore, existing research shows that not only the presence of a pedagogical agent (whether there is an agent or not) but also the cues of an agent (the characteristics of an agent), such as appearance, politeness, feedback, behavior, emotions, and personality, can evoke a positive effect<sup>8,14-17</sup>. Inspired by the persona effect rule, we expect positive results to extend to tedious tasks. In this study, we investigate this issue using virtual agents and a tedious task to induce positive outcomes in human-agent interactions. In particular, we focus on human likeness and responsiveness, which are important attributes of agents studied in previous works; that is, they can activate human social cognition, feelings, and behavioral responses to agents during interaction<sup>18-21</sup>. Especially



**Figure 1.** Our prototype system is in which participants experience a tedious task with a virtual partner (i.e., an agent) in an immersive virtual environment. (a) A user is playing with the prototype system. (b) A virtual agent is designed as a partner. See accompanying demo video for more details.

for responsiveness, we think it is important to give agents the ability to show social responsiveness, and make them adopt proper strategies to respond to humans, which may be beneficial to human-agent interaction. However, their effects on human-agent relationships and performance remain unclear, especially in tedious activities. Therefore, we focused on human likeness and responsiveness as key social cues of agents and studied their effects on the human-agent relationship and performance.

In this study, we designed a tedious task by simulating a parcel-sorting assembly line in an immersive virtual environment (IVE). We used this task to build an experimental environment in which user experience and performance in a tedious task were explored (Figure 1). In the designed experimental environment, an agent was used as the participant's virtual partner. In particular, by controlling the agent's appearance, verbal features, and intelligent feedback, the effects of two social cues of the agent were explored: human-likeness (how much an agent resembles a human) and responsiveness (how an agent responds to humans). Regarding responsiveness, we not only examined the effect of its presence (Experiment 1) but also considered the effect of different social response strategies (Experiment 2). This study systematically examined the potential effects of these cues on the human-agent relationship and user performance in a tedious task. The contributions of this study are as follows.

- An elaborate VR environment is designed to study user experience in tedious tasks through human-agent relationships.
- The effects of an agent's human likeness and responsiveness on human-agent relationships and user performance in a tedious task are examined in the VR environment.
- The effects of an agent's human-likeness and social response strategy on human-agent relationships and user performance in a tedious task were examined.

The remainder of the paper is organized as follows. We briefly review related work in Section [Related works](#). Then we present our developed tedious task in an IVE in Section [Tedious Task Construction by Sorting Parcels on the Virtual Assembly Line](#) and present two empirical experiments in Sections [Experiment 1: Effect of Agent's Human-likeness and Responsiveness in Tedious tasks](#) and [Experiment 2: Exploring the Effect of Social Response Strategy in Agent's Responsiveness](#), respectively. Finally, we discuss main findings of our studies in Section [General Discussion](#) and conclude our work in Section [Conclusions](#).

## Related works

### Human-agent relationship

Intelligent agents are becoming increasingly widespread, which promotes the study of human-agent interaction<sup>2</sup>. Studies have proposed that effective human-agent interaction promotes positive motivational, behavioral, and cognitive outcomes for increased task performance efficiency<sup>8,22,23</sup>. The positive outcomes of human-agent relationships have also been concerned.

The human-agent relationship refers to the social aspect of human-agent interactions. Few studies have investigated the relationship between human users and several types of computer agents. In early studies of human-agent relationships, Bickmore et al. focused on agents designed to establish and maintain long-term social-emotional relationships with human users<sup>24</sup>. They found that compared to agents without any deliberately designed social-emotional or relationship-building abilities, agents with these abilities obtain more respect, likeness, and trust from human participants. Recent studies have

explored methods to improve an agent's ability to establish relationships with users and further confirm the effectiveness of this ability<sup>5,25</sup>. Therefore, improving human-agent relationships can improve human experience and performance in human-agent interactions, leading to positive changes in behavioral, cognitive, or emotional states.

### **Study in tedious activities**

People often have to engage in tedious activities in various scenarios, and many of these activities are related to interactions, such as selecting targets in visual navigation<sup>26</sup>. Tedious tasks easily lead to fatigue and poor user experience<sup>9-11</sup> and even reduce people's motivation (to participate in activities) and performance. Interactive learning activities can prove ineffective if the tasks are tedious and time-consuming, because fatigue and frustration may negate the positive learning effect, leading to poor motivation in subsequent sessions<sup>27</sup>.

One way to address tedious tasks is to automate and intelligently design technologies to replace manual labor. For example, manually detecting depression is a time-consuming and tedious task. Here, Ay et al. developed a fully automated depression diagnosis system by analyzing electroencephalogram (EEG) signals<sup>28</sup>. Other studies addressed tedious work by optimizing the designs<sup>27,29</sup>. An example presented in<sup>27</sup> was that Tom Sawyer introduced a design concept that transforms image labeling and other tedious tasks into entertainment through redesigning tasks. However, in many practical situations, automated methods are difficult or even disallowed, and the cost of redesigning tasks is high. It is essential to find effective methods to enhance the user experience without altering the primary tasks. In this study, we explore these methods by establishing a positive human-agent relationship in a tedious task.

### **Persona Effect and Social Cues of Agent**

Using agents in a virtual environment is a typical element and an added value that has a significant impact on user experience<sup>4</sup>. The positive effect of virtual agents occurring in an interactive virtual learning environment has been studied and is known as the persona effect. Persona Effect is defined as that *the presence of a life-like character in an interactive learning environment — even if they are not expressive — can have a strong positive impact on student's perception of their learning experience*<sup>13</sup>.

Designing appropriate cues (especially social cues) for agents helps to induce the persona effect. Previous studies demonstrated that agents' social cues (for example, human likeness, role, dress, sociality, emotion, and personality) have positive effects on learners' mental and behavioral outcomes, such as social judgements, interest to learn, self-efficacy, and learning performance<sup>7,8,15,16,30,31</sup>.

To interpret the persona effect, the *Media Equation* theory proposed by Reeves et al. was commonly used, which states that people exhibit remarkable social reactions to computers and other media (including agents) and treat them as real people, even if they have actual real feelings, intentions or human motivations<sup>32</sup>. When a computer or robot presents a set of cues (such as language and behaviors) that would normally be associated with humans, humans respond by exhibiting social behaviors and making social attributions<sup>33</sup>. Therefore, human-agent interaction reflects social cues prevalent in human-human interaction, and positive effects in human-human interaction can occur to a certain extent<sup>22,30</sup>.

Inspired by the persona effect, we expect the positive effects of an agent to sustainably exist in human-agent relationships. For these agents to function properly, they require social cues to exhibit a certain amount of social skills or sociality<sup>5,34</sup>. As mentioned above, human likeness and responsiveness are two important attributes of agents that have been studied previously. Human-likeness is a social cue for a human's social response to an agent or a robot<sup>18</sup>. A key characteristic of human likeness is appearance. The appearance reflects the shape of the body and the degree to which it resembles the human body<sup>35,36</sup>. Responsiveness is another social cue that facilitates interpersonal interaction<sup>20</sup>. These cues can activate human social cognition, feelings, and behavioral responses to agents during interaction<sup>18-21</sup>.

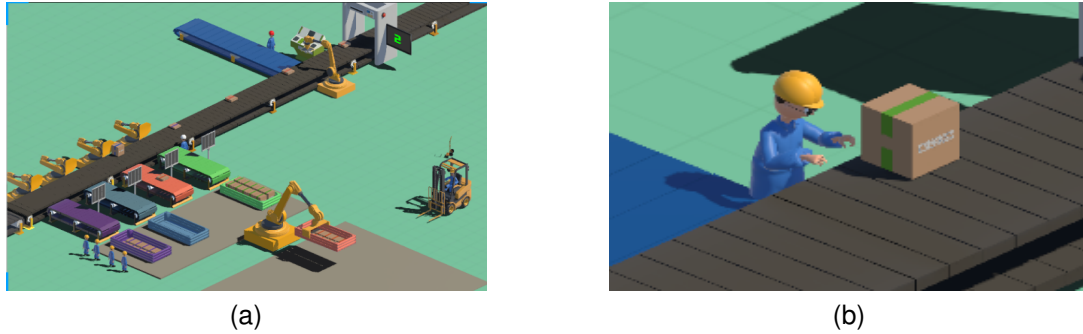
Based on previous studies, human likeness and responsiveness are expected to play a positive role in human-agent relationships and performance in human-agent interaction tasks. However, their impact on these aspects remains unclear, especially in tedious activities. Therefore, this study considers human likeness and responsiveness as two key social cues of agents to study their effects on human-agent relationships and performance by simulating parcel sorting on an assembly line in an immersive virtual environment (IVE) as a tedious task.

## **Tedious Task Construction by Sorting Parcels on the Virtual Assembly Line**

Several tedious tasks have been described in previous studies. However, to the best of our knowledge, there is still a lack of a typical tedious task paradigm or platform for user experience studies. In this study, we constructed a tedious task in an IVE. Details are as follows.

### **Task and design scenario**

To identify an appropriate scenario for a tedious task, we conducted brief interviews with five college students followed by a literature review. In the interview, interviewees mentioned that they had once worked as interns in a nearby factory during



**Figure 2.** Our prototype system is in which participants experience a tedious task with a virtual partner (i.e., an agent) in an IVE. (a) A user is playing with the prototype system. (b) A virtual agent is designed as a partner. See accompanying demo video for more details.

summer vacation, and they all reported that sorting items repeatedly for a long time on an assembly line is a tedious task. One interviewee stated that *some of my classmates and I once went to a factory to do an internship on the assembly line. One week later, we all quit. The work was simple but extremely time-consuming, repetitive, and tedious.* Our literature review also showed that tasks on an assembly line are generally tedious, repetitive, and time-consuming<sup>37</sup>.

Furthermore, we found that using a tedious task on the assembly line is suitable for setting up agents with controllable cues; thus, we used this scenario for our study.

## Construction of the prototype system

### Software of the system

A prototype system of a virtual assembly line for parcel sorting was developed using the Unity 3D 2019 software. Two snapshots of the virtual assembly line are shown in Figure 2.

*Task.* The user was designed to be an intern who participated in the internship. During the internship, the user was required to manually complete the task of sorting parcels (Figure 2b). Many parcels (including five types) were transported at a certain frequency on the assembly line. The parcel can be successfully sorted only by pressing the corresponding number keys in a timely and accurate manner. The user's performance was displayed on the screen in the IVE in real time.

*Events.* In addition to sorting parcels, the user must deal with several events during the internship experience. These events were designed to simulate real-world situations encountered during daily activities. Users were instructed to select an action from the preset options when addressing these events.

*Agent.* The agent's role was designed as that of the user's partner. The agent was not designed to help users complete the task directly but to provide suggestions and feedback based on the user's behavior. The appearance, voice, and feedback of the agent can be manipulated flexibly.

### Hardware of the system

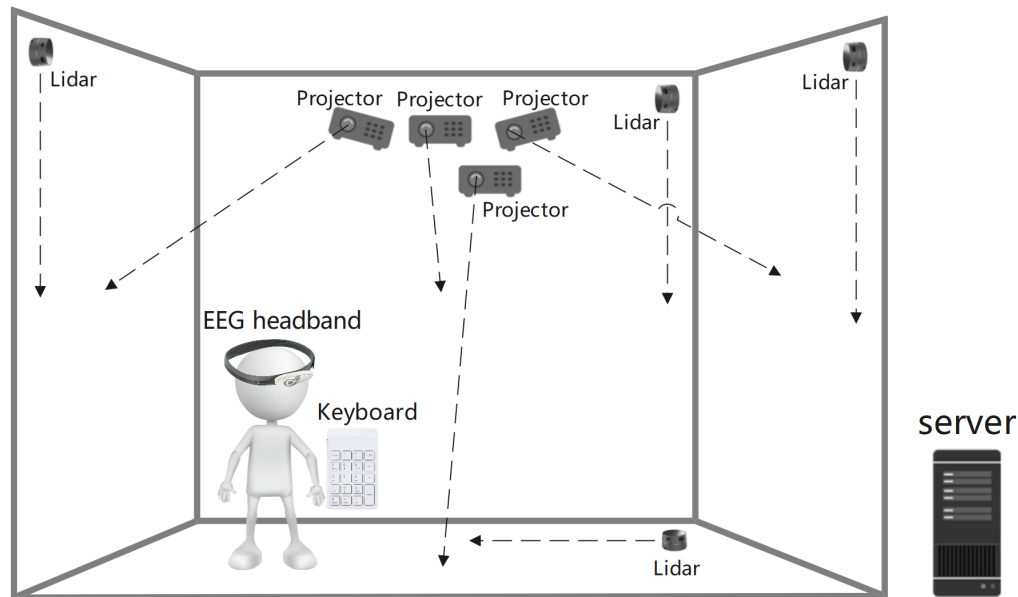
To ensure a highly immersive experience, the system was built into an interactive IVE. The overall framework of the system hardware is shown in Figure 3. The details of the components are described below.

*Display devices.* The system was displayed in a cave automatic virtual environment (CAVE)<sup>38</sup>, where virtual contents were projected onto four surfaces (three walls and the floor) of a room-sized cube to represent a specific pre-designed IVE.

*Interactive devices.* Each of the four surfaces projected in the CAVE was equipped with a 360° laser range finder, YDLIDAR G2. These laser range finders scan at 5–12 Hz and range from 0.1–12 m, enabling multitouch interaction on all four surfaces. During the task, the user selects actions by directly touching virtual content on the surface. The touch position was identified by using laser rangefinders.

The system provides two interactive ways to complete the parcel-sorting task. One way is to use a Bluetooth keyboard to ensure that users can press numbers directly (Figure 4a). The other method is using an RGB camera with the aid of machine vision algorithms to identify the number on the physical cardboard that the user has picked up (Figure 4b). The latter method can provide a higher level of physical challenges than the former<sup>39</sup>.

*Attention evaluation device.* We used a portable EEG headband (Brainlink Pro) equipped with a NeuroSky ThinkGear sensor<sup>40</sup> to assess the sustained attention level of users. It has received authorization from the U.S. Federal Communications Commission and has been approved for use in the European Union<sup>41</sup>. The NeuroSky ThinkGear is a convenient non-invasive biosensor that has been widely used in brain-computer interface systems<sup>42,43</sup>. It connects to three dry EEG electrodes and



**Figure 3.** Overall framework of the system hardware.

comprises three dry EEG electrodes (i.e., active, reference, and ground), placed on the forehead using an elastic band. The dry EEG electrodes were placed close to the standard EEG locations F7, Fp1 (bipolar channel), and Fpz (ground electrode) according to the International 10–20 system. The NeuroSky sensor samples EEG activity at frequencies up to 512 Hz. We used the headband to monitor the user’s EEG signal in real-time and transmit the EEG data to the server via a wireless Bluetooth connection<sup>40</sup>.

The overall environment of the prototype system is shown in Figure 1a. Based on the constructed prototype system, we conducted two experiments (presented in Sections [Experiment 1: Effect of Agent’s Human-likeness and Responsiveness in Tedious tasks](#) and [Experiment 2: Exploring the Effect of Social Response Strategy in Agent’s Responsiveness](#)) to explore the effects of an agent’s human-likeness and responsiveness on human-agent relationships and performance in the parcel sorting task.

### Usability of the system

To evaluate the performance of our designed system in terms of usability and immersion, we conducted a test using a sample of eight users. Usability was quantified using the system usability scale (SUS), which can identify systems with good/poor usability<sup>44</sup>. Immersion was measured using the immersive experience questionnaire (IEQ), which has shown good reliability and has been widely used in previous VR studies<sup>45,46</sup>. The results indicate that the total SUS score is  $89.06 \pm 5.28$ , which means the system is acceptable in terms of acceptability (above 70), and excellent in terms of adjective rating (above 85), suggesting good usability of our system. The immersion scores were then calculated by summing all questions in the IEQ. The IEQ score of our system was  $120.75 \pm 6.86$ , which is higher than that of the highly immersive VR system in a previous study ( $118.3 \pm 11.3$ <sup>46</sup>). These results suggest that the proposed system can provide users with good usability and immersion.

## Experiment 1: Effect of Agent’s Human-likeness and Responsiveness in Tedious tasks

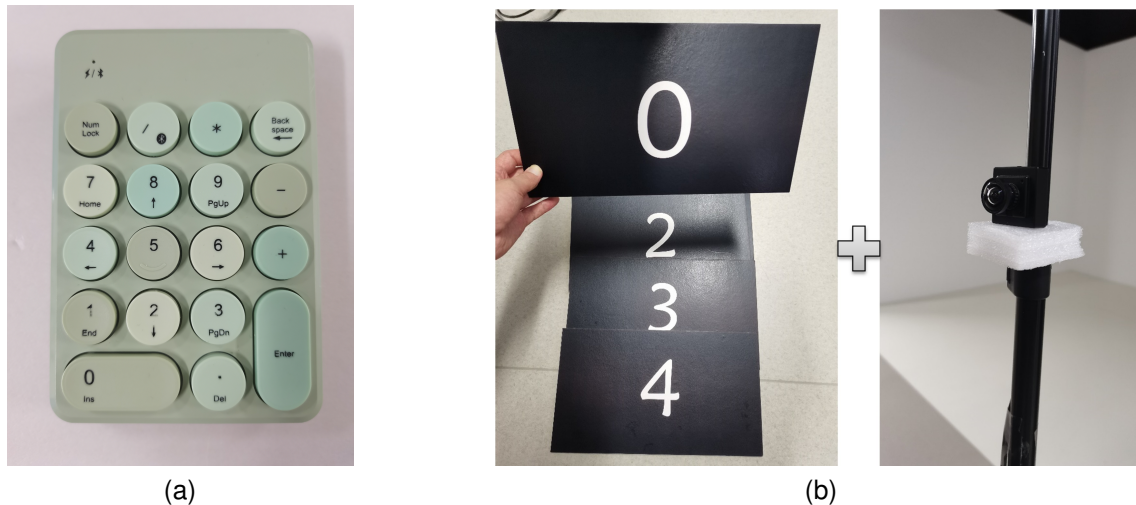
### Purpose

In this section, we experimented to explore the effect of an agent’s human-likeness and responsiveness on the human-agent relationship and performance in a tedious task, that is, the parcel sorting task summarized in Section [Tedious Task Construction by Sorting Parcels on the Virtual Assembly Line](#).

### Participants

Sixteen undergraduate students (10 males and 6 females) were recruited for this study. Their ages ranged from 19 to 25 years, with an average of 20.31 years ( $SD = 1.49$  years). All patients had normal hearing and normal or corrected-to-normal vision. Informed consent was obtained from all the participants. This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Human Experimentation at Jining No.1 People’s Hospital.





**Figure 4.** Some key components are used in the system hardware. (a) Bluetooth keyboard. (b) Physical boards (with five numbers) and an RGB camera.

### Experimental design

Our experiment used a  $2$  (human-likeness of agent: robot-like/ human-like)  $\times$   $2$  (responsiveness of agent: with responsiveness/without responsiveness) factorial, within-subject design. Considering that human likeness might significantly affect the human-agent relationship<sup>47,48</sup>, we designed two agents with different human likeness by controlling their appearance and voice:

- Robot-like agent: This virtual agent was designed as the participant’s companion in the form of a robot (Figure 5a) with a robot-style voice.
- Human-like agent: This virtual agent was designed to be the participant’s companion with a humanoid appearance (Figure 5b) and a natural human voice.

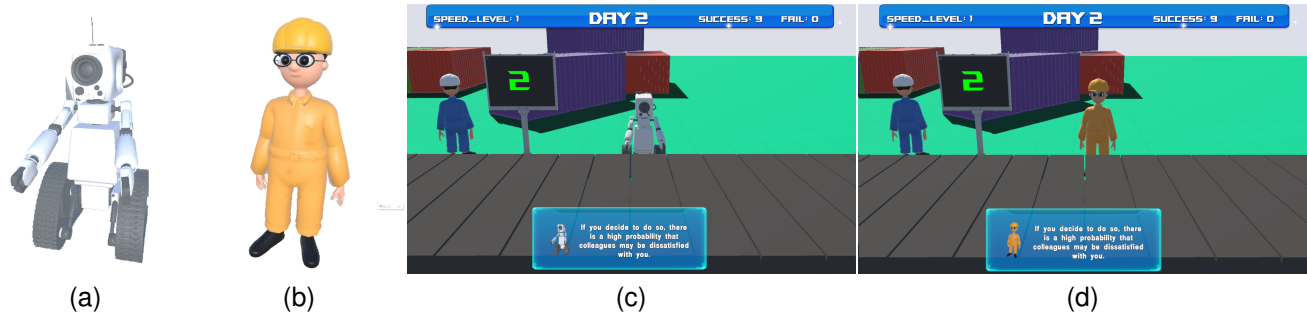
We designed the responsiveness of agents by controlling their interactive feedback to participants:

- Agent with responsiveness: The agent can actively provide reminders and suggestions in voice or text according to the participant’s operations during the task (Figure 5c and 5d).
- Agent without responsiveness: The agent did not provide reminders or suggestions to the participants.

Given the  $2 \times 2$  factorial within-subject design, there were four types of agents:

- Virtual robot-like agent without responsiveness
- Virtual robot-like agent with responsiveness
- Virtual human-like agent without responsiveness
- Virtual human-like agent with responsiveness

Both factors (human-likeness and responsiveness) of the agents were within-subject factors, and each participant needed to experience all four types. Furthermore, the individual characteristics of the participants remained unchanged in the within-subject design. To avoid the order effect, the order of experience of the four types was counterbalanced using a Latin square design. There were four sequences, and accordingly, the number of participants was a multiple of four. The dependent variables in the design consisted of two aspects: perceived human-agent relationship (perceived closeness, intimacy, and involvement with the agent) and task performance (the accuracy of the task and the attention level), which were analyzed below.



**Figure 5.** Design of human-likeness and responsiveness of agent in Experiment 1. (a) The robot-like virtual agent. (b) The human-like virtual agent. (c) A snapshot of text suggestions given by a robot-like agent. (d) A snapshot of text suggestions given by a human-like agent. See accompanying demo video for animation details.

### Apparatus and task

We used the system described in Section [Construction of the prototype system](#) as the experimental environment (Figure 1a). Note that to identify the parcel type, using physical cardboard selected by participants can simulate the interaction in real sorting to a certain extent, which is physically challenging (i.e., less tedious). To increase the tediousness of the task, we had participants use a Bluetooth keyboard instead of a physical cardboard to complete the task.

The virtual assembly line was projected onto the three vertical screens of the CAVE. In the experimental task of this study, the participants were required to sort 72 parcels (including five types marked with five colors). Parcels were randomly generated every 6 s and transported on the assembly line from left to right. When the package was transported to the middle screen of the CAVE, participants were required to sort the package. The parcel can be successfully sorted only by pressing the corresponding number keys in a timely and accurate manner. After each sorting, the system would timely feedback on "correct" or "wrong" through a notice board in the scene.

The agents had two types of interactions with the user: 1) the agents provided voice feedback according to the dynamic task performance of the user. Specifically, if the user made a mistake in the sorting task, the agent will immediately remind the user "We just made a mistake. Next, we must be more careful." If the user continuously and correctly sorts six times, the agent will immediately encourage the user "Well done! We have become increasingly skilled. Keep going." 2) The agents provide voice and text feedback according to the user's choice to address the events. Only the agent partner (robot or colleague dressed in yellow) provides the above feedback, while other agents (dressed in blue) do not provide feedback to the user. The specific feedback is presented in Table 1 of the Appendix.

Moreover, the participants were required to handle three events during the task. When each event occurred, the system simulated and showed the virtual scenarios to the participants, and the participants were required to make a behavioral choice from the alternatives. Meanwhile, the process of sorting parcels is paused. For example, one of the events is that "Consider leaving work early." A colleague offers the participant to drop off unfinished work and leave early to have fun together—this is the scenario of the event. Subsequently, the participant needed to address this event by choosing an action from alternative options, for example, "Give a stern rebuff." When the participants finished the choice, the agent provided corresponding comments and suggestions on the possible consequences of this decision, for example, "Your work principle is correct, but it may be better if you attempt to use a gentler way first." The details of these three events are provided in the Appendix.

### Measures

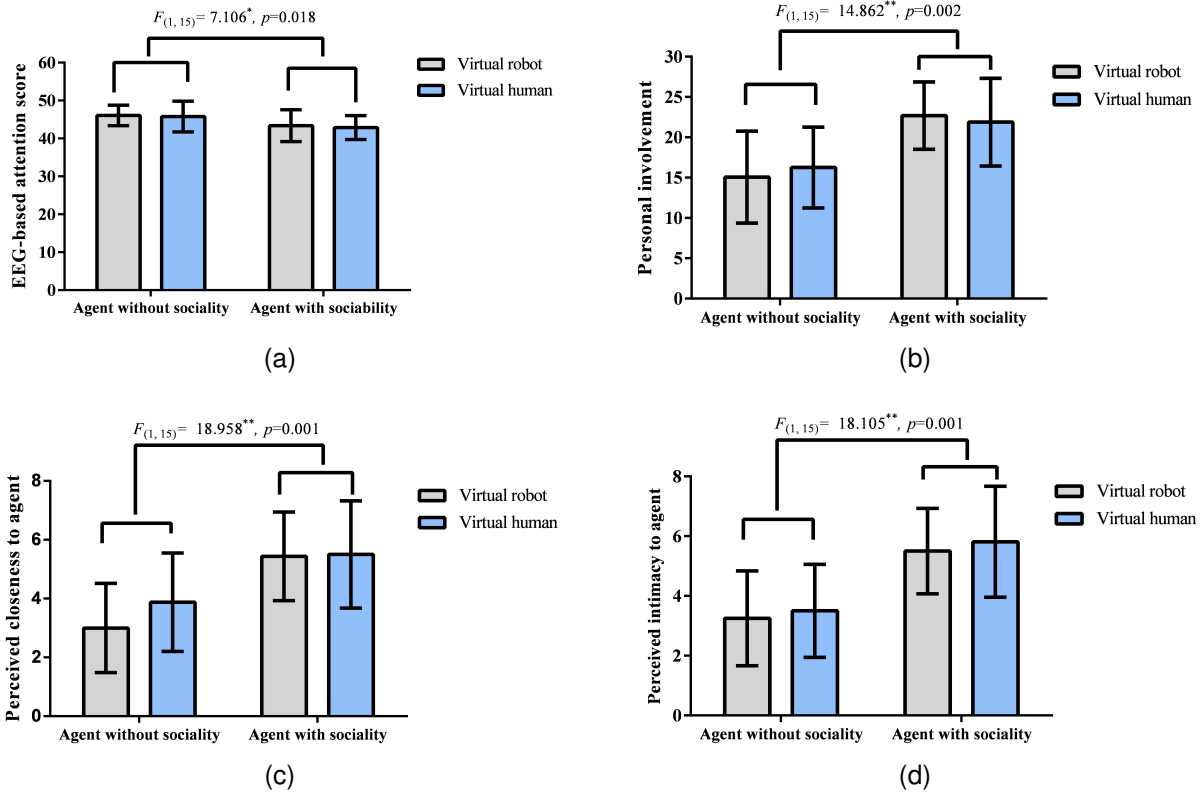
The dependent variables consisted of task performance and the perceived human-agent relationship of the participants. The measures of task performance included the accuracy of the task and attention level during the task. The measures of perceived human-agent relationships include perceived involvement, closeness, and intimacy with agents.

#### Accuracy of task

The system automatically recorded the accuracy of each participant's task. The recorded data included the number of misclassified and correctly classified parcels in the sorting process, and the sorting accuracy rate was calculated to evaluate the task accuracy.

#### EEG-based attention

Attention refers to the intensity of a user's level of mental "focus," such as that which occurs during intense concentration and directed (but stable) mental activity. The proprietary eSense algorithm built into the EEG headband was used to evaluate the



**Figure 6.** Results on participants' attention level reported from EEG device (a), personal involvement to agent (b), perceived closeness (c), and intimacy (d) in Experiment 1. Participants had a significantly lower level of attention on agents with responsiveness than that without responsiveness ( $*p < 0.05$ ,  $**p < 0.01$ , confidence intervals(CIs) in figures represent 95% CIs). However, participants perceived significantly higher level of involvement, closeness, and intimacy in agents with responsiveness than those in agents without responsiveness.

player's attention level based on real-time measured EEG signals. First, the collected EEG signal is pre-processed to remove eye movement artifacts. Then, the eSense algorithm works by performing a Fast Fourier Transform on the collected EEG data, and then the EEG wave are filtered to obtain alpha and beta waves according to the requirements. The power spectrum is analyzed and the basic reference value is selected to complete the normalization. Finally, Attention level is calculated according to the power spectrum of EEG band. That is, the eSense algorithm extracts metrics related to attention, and output attention values (ranging from 0 to 100) at 1 Hz, indicating the intensity of concentration on selective information (whether subjective or objective)<sup>49,50</sup>. The attention values reported by the eSense algorithm have been demonstrated to have strong positive correlations with self-assessment and other physiological measures of attention<sup>42,43,51</sup>. Distractions, mind-wandering, inattention, and anxiety can reduce attention levels. The attention value of each player was recorded during the task.

### Personal involvement

Personal involvement is operationally defined as important to the person and can be activated by the presence of situational and/or intrinsic self-relevance to an object<sup>52</sup>. This is an important measure of relationships. In our study, involvement refers to participants' involvement with agents. Thus, involvement is operationalized as the perceived importance and/or self-relevance of virtual agents to the participant. Personal involvement was measured using a 5-item scale adapted from Novak et al.'s work<sup>53</sup> by Liu et al<sup>52</sup>. We adapted the scale by adding the context "How do you feel about the agent?" to make the items suitable for measuring personal involvement with agents. All items adopted a seven-point Likert scale, with anchors ranging from 1 (total disagreement) to 7 (total agreement). The reliability of the scale is good ( $\alpha = 0.845$ ).

### Perceived closeness and intimacy in relationship

A sense of closeness and intimacy is essential for building and maintaining relationships, achieved through social interaction<sup>54,55</sup>. Most existing measures of closeness and intimacy were designed to measure daily relationships with humans, such as the emotional closeness questionnaire (ECQ), relational intimacy questionnaire (RIQ)<sup>56</sup> and Flores et al.'s 5-item questionnaire<sup>57</sup>.



They are unsuitable for measuring the current situational experiences with agents. In our study, we used the measure of emotional closeness from China Family Panel Studies (CFPS)<sup>58</sup>. This measure uses a single item on an 11-point scale ranging from 0 (not close at all) to 10 (extremely close) to measure closeness. A single-question measure is commonly used to measure closeness<sup>59</sup>, which can avoid confusion when participants understand an item. Similarly, we used a single item on an 11-point scale to measure perceived intimacy.

## Procedure

At the beginning of the experiment, the participants were given a detailed introduction to the task and how to operate the system. Subsequently, each participant experienced all four types of agents completing the task in a certain order. Each round of the task took approximately 12 min. The attention-level output of the EEG device was recorded. When each type of task was completed, the participants filled out a questionnaire containing the scales in the Sections [Perceived closeness and intimacy in relationship](#) and [Personal involvement](#). After the experiment, each participant answered a free-response question to share their experience.

## Results

First, Kolmogorov–Smirnov tests (K-S tests) were performed. Although the sample size was small, the results support the normality assumption of normality ( $ps > 0.05$ ). We then performed numerous point-biserial correlations, and the results did not show significant correlations between sex and dependent variables ( $ps > 0.05$ ). To explore the effects of an agent’s human likeness and responsiveness on task performance, two  $2 \times 2$  repeated-measures ANOVAs were performed with the accuracy of the task and the EEG-based attention level as independent variables. To explore the effects of an agent’s human likeness and responsiveness on the experience of human-agent relationships, three  $2 \times 2$  repeated measures ANOVAs were performed with personal involvement, perceived closeness, and intimacy with agents as independent variables. The results are as follows.

### Results on the accuracy of task

The results for the accuracy of the task did not reveal any main effects. Neither human-likeness [ $F_{(1,15)} = 1.756, p = 0.205, \eta_p^2 = 0.105$ ] nor responsiveness [ $F_{(1,15)} = 0.850, p = 0.371, \eta_p^2 = 0.054$ ] significantly affected the accuracy. The interaction effect of human-likeness  $\times$  responsiveness was not significant [ $F_{(1,15)} = 3.768, p = 0.071, \eta_p^2 = 0.201$ ].

### Results on the attention

The results for the EEG-based attention level revealed a significant main effect of responsiveness [ $F_{(1,15)} = 7.106, p = 0.018, \eta_p^2 = 0.321$ ] (Figure 6a in Experiment1). Participants had a higher level of attention to agents without responsiveness ( $M = 45.94, SD = 6.35$ ) than to agents with responsiveness ( $M = 43.13, SD = 6.83$ ).

Neither the main effect of agent’s human-likeness [ $F_{(1,15)} = 0.066, p = 0.800, \eta_p^2 = 0.004$ ] nor the interaction effect [ $F_{(1,15)} = 0.003, p = 0.096, \eta_p^2 = 0.000$ ] was significant.

### Results on personal involvement

The results on personal involvement also indicated a significant main effect of responsiveness [ $F_{(1,15)} = 14.862, p = 0.002, \eta_p^2 = 0.498$ ] (Figure 6b in Experiment1). Participants perceived a higher level of involvement with agents with responsiveness ( $M = 22.28, SD = 8.97$ ) than with agents without responsiveness ( $M = 15.66, SD = 9.93$ ).

Neither the main effect of the agent’s human-likeness [ $F_{(1,15)} = 0.043, p = 0.838, \eta_p^2 = 0.003$ ] nor the interaction effect [ $F_{(1,15)} = 1.132, p = 0.304, \eta_p^2 = 0.070$ ] was significant.

### Results on perceived closeness

The results on perceived closeness revealed a significant main effect of responsiveness [ $F_{(1,15)} = 18.958, p = 0.001, \eta_p^2 = 0.558$ ] (Figure 6c in Experiment1). Specifically, participants perceived a higher level of closeness to agents with responsiveness ( $M = 5.47, SD = 3.09$ ) than agents without responsiveness ( $M = 3.44, SD = 2.98$ ).

Neither the main effect of the agent’s human-likeness [ $F_{(1,15)} = 0.929, p = 0.350, \eta_p^2 = 0.058$ ] nor the interaction effect of human-likeness  $\times$  responsiveness was significant [ $F_{(1,15)} = 1.489, p = 0.241, \eta_p^2 = 0.090$ ].

### Results on perceived intimacy

Similar to the results on perceived closeness, the results on perceived intimacy revealed a significant main effect of responsiveness [ $F_{(1,15)} = 18.105, p = 0.001, \eta_p^2 = 0.547$ ] (Figure 6d in Experiment1). Specifically, participants perceived a higher level of intimacy with agents with responsiveness ( $M = 5.66, SD = 3.06$ ) than those without responsiveness ( $M = 3.38, SD = 2.90$ ).

Neither the main effect of agent’s human-likeness [ $F_{(1,15)} = 0.396, p = 0.539, \eta_p^2 = 0.026$ ] nor the interaction effect of social role  $\times$  responsiveness [ $F_{(1,15)} = 0.011, p = 0.920, \eta_p^2 = 0.001$ ] was significant.

### Results of the free-response question

The results indicated that 87.5 % of participants shared their feelings, and all of them used tedious related words (for example, tedious, boring, time-consuming, repetitive, monotonous et al. see Table 1) to describe their feelings regarding the task; the remaining 12.5% did not share their feelings. These results supported the idea that participants felt tedious about the task.

**Table 1.** Open coding feelings and frequency for each feeling after Experiment 1

Open coding results (feelings)	Frequency	Open coding results (feelings)	Frequency
boring	16	tedious	5
repetitive	6	time-consuming	6
monotonous	5	sleepy	3
uninteresting	4	too easy/simple	2
focused	3	calm	2

### Brief Discussion of Experiment 1

Based on the above results, we conclude that responsiveness is an important cue for agents to improve human-agent relationships. Despite the responsiveness of the agent increased, this did not impact the accuracy of the task, showing that the participants had redundant attention while doing the tedious task. Responsiveness decreased the participants' attention because they attracted redundant attention to their kind reminders and suggestions. This would not decrease the task performance of participants but helped them develop a positive relationship with agents while working with them. Participants perceived a significantly higher level of involvement with agents with responsiveness, indicating that they felt this type of agent was more important and self-relevant<sup>52</sup>. Accordingly, they felt the agents with responsiveness were more like partners who tried to engage in the task and provided some help, and this characteristic made the participants feel significantly higher closeness and intimacy with those agents.

Based on the effect of responsiveness, we further considered that not only was the responsiveness of agents important in tedious work, but also the strategy to show the social response (i.e., social response strategy) to humans might be important. Moreover, there were no significant main effects of the agent's human likeness, which should be further examined. To explore this point, we further focused on the social response strategy and presented the next experiment in Section [Experiment 2: Exploring the Effect of Social Response Strategy in Agent's Responsiveness](#).

## Experiment 2: Exploring the Effect of Social Response Strategy in Agent's Responsiveness

### Purpose

In Experiment 1, the agent's responsiveness relied on whether the agent provided verbal reminders and suggestions. In this section, we present an experiment that further explores and decomposes this condition into two social response strategies. The agent provides reminders and suggestions when participants make a positive or negative decision. We explore the effect of these two strategies on the human-agent relationship and performance in a tedious task.

### Participants

Another 16 undergraduates, including 12 men and 4 women, who did not participate in Experiment 1 were recruited. Their ages ranged from 19 to 28 years, with an average of 23.31 years ( $SD = 2.03$  years). All patients had normal hearing and normal or corrected-to-normal vision. Informed consent was obtained from all the participants. This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Human Experimentation at Jining No.1 People's Hospital.

### Experimental design

Our experiment used a 2 (human-likeness of agent: robot-like, human-like)  $\times$  2 (social response strategies: responding to positive/negative decision) factorial within-subject design. As in Experiment 1, both cues were within-subjects factors. The condition of human likeness was the same as in Experiment 1, while responsiveness was further divided into two conditions according to how the agents responded to the participants. As presented in the section [Software of the system](#), participants needed to deal with three events during the internship experience. For each event, participants were presented with different options in the system; these options had different appropriateness for dealing with the events. Making a positive decision meant that participants selected the appropriate option while making a negative decision meant that participants selected the inappropriate option<sup>60</sup>. When experiencing the two social response strategies, the participants were instructed to respond negatively or positively to induce agent responses, and then their perception of the human-agent relationship was measured. In

this experiment, we must make the participants conscious of what they have made as a positive/negative choice; thus, it is a better way to make positive and negative responses according to their own understanding.

Specifically, two social response strategies were set as follows:

- Response to positive decision. Participants were told to “respond positively” according to their own understanding. When they made a positive decision, the agent responded by providing comments and suggestions. For example, given the event that a virtual colleague comes to the participant and offers to drop off unfinished work and leave early to have fun together, the participant addresses this by choosing the option of “Decline politely.” The agent then provided comments and suggestions for this decision.
- Response to negative decision. Participants were told to “respond negatively” according to their own understanding. When they made a negative decision, the agent responded by providing comments and suggestions. Following the above example, given the same event, the participant deals with this event by choosing the option of “Accept the colleague’s proposal.” The agent then provided comments and suggestions for this decision.

Given the  $2 \times 2$  factorial within-subject design, there were four types of agents:

- Virtual robot-like agent with the strategy of responding to positive decisions.
- Virtual robot-like agent with the strategy of responding to negative decisions.
- Virtual human-like agent with the strategy of responding to positive decisions.
- Virtual human-like agent with the strategy of responding to negative decisions.

All the participants were required to experience all four types. To avoid the order effect, the order of experience of the four types was counterbalanced with a Latin square design. There were four sequences, and accordingly, the number of participants was a multiple of four. The dependent variables in the design consisted of two aspects: perceived human-agent relationship (perceived closeness, intimacy, and involvement with the agent) and task performance (the accuracy of the task and the attention level). These were similar to those in Experiment 1.

### **Apparatus, environment, measures and procedure**

The apparatus, experimental environment, measures of dependent variables, and experimental procedure were similar to those in Experiment 1 (Section [Experiment 1: Effect of Agent’s Human-likeness and Responsiveness in Tedious tasks](#)).

### **Results**

First, K-S tests were performed, and the results supported the assumption of normality ( $ps > 0.05$ ). As in Experiment 1, the results of point-biserial correlation did not reveal any significant correlation between gender and each dependent variable ( $ps > 0.05$ ), suggesting that gender is not a factor that significantly affects the results. To explore the effects of an agent’s human-likeness and social response strategy on task performance, two  $2 \times 2$  repeated-measures ANOVAs were performed with the accuracy of the task and the EEG-based attention level as independent variables. Subsequently, to explore the effects of an agent’s human-likeness and social response strategy on the experience of the human-agent relationship, three  $2 \times 2$  repeated measures ANOVAs were performed with personal involvement, perceived closeness, and intimacy (to agents) as dependent variables. The results are as follows.

#### **Results on the accuracy of task**

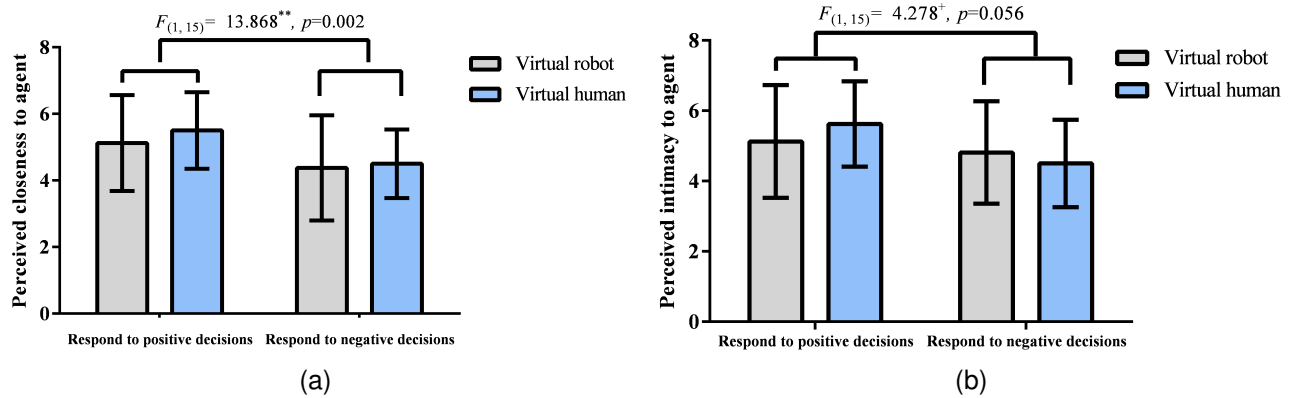
The results for the accuracy of the task did not reveal any significant effects. None of the agents’ human likeness, social response strategy, or interaction effect significantly affected the performance ( $ps > 0.05$ ).

#### **Results on the attention**

The results for the EEG-based attention level did not reveal any main effects. Neither social response strategy [ $F_{(1,15)} = 4.093$ ,  $p = 0.061$ ,  $\eta_p^2 = 0.214$ ], nor human-likeness [ $F_{(1,15)} = 4.032$ ,  $p = 0.063$ ,  $\eta_p^2 = 0.212$ ] showed a significant effect. The interaction effect of human-likeness  $\times$  social response strategy was not significant [ $F_{(1,15)} = 0.215$ ,  $p = 0.650$ ,  $\eta_p^2 = 0.014$ ].

#### **Results on personal involvement**

The results for involvement did not reveal any significant effects. None of the agents’ human likeness, social response strategy, or interaction effects significantly affected personal involvement ( $ps > 0.05$ ).



**Figure 7.** Results on perceived closeness (a) and intimacy (b) in Experiment 2. Participants perceived a significantly higher level of closeness to agents responding to positive decisions than agents responding to negative decisions. Moreover, participants had a marginally higher level of intimacy with agents responding to positive decisions than agents responding to negative decisions ( $+0.05 < p < 0.06$ ,  $** p < 0.001$ , CIs in figures represent 95% CIs).

### Results on perceived closeness

The social response strategy had a significant main effect on perceived closeness [ $F_{(1,15)} = 13.868$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.480$ ]. Agents with a strategy of responding to positive decisions led to higher closeness than agents with a strategy of responding to negative decisions (Figure 7a in experiment2). The main effect of human likeness and interaction effect of human-likeness  $\times$  social response strategy was not significant ( $ps > 0.05$ ).

### Results on perceived intimacy

The results for perceived intimacy did not reveal any significant effects. None of the agents' human likeness, social response strategy, or interaction effects significantly affected perceived intimacy ( $ps > 0.05$ ). However, the main effect of the social response strategy could be regarded as marginally significant [ $F_{(1,15)} = 4.278$ ,  $p = 0.056 < 0.06$ ,  $\eta_p^2 = 0.222$ ]. Agents with a strategy of responding to positive decisions led to higher intimacy than agents with a strategy of responding to negative decisions (Figure 7b in experiment2).

### Results of the free-response question

The results showed that 100% of participants shared their feelings, and all of them used tedious-related words (see Table 2) to describe their feelings regarding the task. These results supported the hypothesis that participants felt tedious about the task.

**Table 2.** Open coding feelings and frequency for each feeling after Experiment 2

Open coding results (feelings)	Frequency	Open coding results (feelings)	Frequency
boring	17	tedious	11
repetitive	5	time-consuming	6
monotonous	3	sleepy	2
uninteresting	2	too easy/simple	2
cared	2	calm	3

## Brief Discussion of Experiment 2

The results of this experiment showed that the social response strategies of the agent were also important cues that significantly affected perceived closeness and marginally affected perceived intimacy without interfering with task performance.

Recent research has shown that highly social agents make users feel more comfortable and less anxious about building interpersonal closeness with them<sup>55</sup>. Agents using a positive social response strategy make participants feel more positive features (for example., more comfortable and less anxious), leading to a significantly higher level of closeness and a marginally higher level of intimacy. Similar to the results in Experiment 1, the agent's human likeness had no significant effect on the human-agent relationship and task performance. This further supports the result that agents' vocal feedback ability had a stronger impact on human perception than differences in appearance.

## General Discussion

In this study, we investigate the human-agent relationship and task performance in a tedious task when humans interact with virtual agents that have different social cues. We have discussed the results of Experiments 1 and 2 in Sections [Brief Discussion of Experiment 1](#) and [Brief Discussion of Experiment 2](#), respectively. In this section, we present a general discussion of the results of both experiments.

### Attention in Tedious Task

The results showed that the responsiveness of agents has a significantly negative effect on attention, but has no effect on task accuracy, and has a significantly positive effect on the human-agent relationship.

Tedious tasks are often repetitive and easy to perform. Although certain types of tedious tasks require undistracted attention and are sensitive to relatively brief lapses in attention (such as manual inspection of quality control data<sup>61</sup>), other tedious tasks allow people to allocate their attention<sup>62</sup>. From the perspective of cognitive resources, the participants did not need to spend all their cognitive resources on completing the tedious task<sup>63</sup>. Research on robot agents has shown that they can attract the attention of people by speaking<sup>34</sup>. Therefore, when the agent showed responsiveness by speaking, the participant's surplus attention was automatically directed to the content of speaking from the ongoing task of sorting parcels. However, reduced attention did not affect the accuracy of tedious tasks.

We believe that the surplus attention allocated to the agents helps participants increase the chance of being involved with them and ultimately leads to positive human-agent relationships, including a better feeling of closeness and intimacy. This feeling of closeness and intimacy with agents can be particularly valuable for people to engage in activities<sup>64</sup>, which are crucial to the development of a long-term relationship (not only short-term effects)<sup>55,65</sup> and motivation to participate in human-agent collaborative tasks, including tedious activities.

### The Effect of Responsiveness

The results showed that both the presence of responsiveness and responsive strategies have an important impact on the human-agent relationship.

Regarding the effect of an agent's responsiveness, previous research has shown that when individuals are faced with certain situations (for example, tedious or stressful situations), only presenting a virtual partner is insufficient; what significantly improves an individual's emotional experience is the explicit demonstration of attentiveness and responsiveness by the partner<sup>34,55</sup>. Providing appropriate verbal/nonverbal feedback from agents helps users feel cared for and understood, leading to a high degree of acceptance of the agent<sup>4,66</sup>. Therefore, agents with responsiveness can facilitate a sense of rapport and personal involvement through which users can more easily feel closeness and intimacy in human-agent interaction<sup>67-69</sup>.

Regarding the effect of different social response strategies on human-agent relationships, agents using the strategy of responding to positive decisions may be perceived with a more positive impression of social perception. Interpersonal feedback with a positive tendency may increase participants' motivation and feelings of competence<sup>12</sup>, making them feel more comfortable and less anxious about building interpersonal closeness with the agents<sup>55</sup>. Consequently, this strategy leads to higher interpersonal closeness and intimacy. By comparison, although agents using the strategy of responding to negative decisions also provide rational suggestions, this might still lead to negative feelings and perceptions of participants, leading to lower closeness and intimacy.

### The Effect of Agent's Human-likeness

It is worth noting that human likeness is not a social cue that significantly affects human-agent relationships and task performance in this study.

Previous studies on social robots have shown that more human-like features (e.g., body shape, face, and voice) make people perceive them as more human-like. This perceived human likeness is an important determinant of human responses to social robots and has been shown to have a positive effect on human-robot interactions<sup>35,36</sup>. Our results are inconsistent with these findings; one possible reason is that the agents in this study are all cartoon style which may weaken the potential effect of human likeness. Another possible reason is that there are no sensitive questionnaires on the agent's appearance included in this paper. Therefore, the effect of human likeness on the human-agent relationship should be further examined in future work.

Previous research on human-robot interaction shows that dialogue ability has a greater impact on humans' mental model than physical appearance<sup>70</sup>, which matches our findings that responsiveness, rather than human-like physical qualities, is key for cultivating human-agent relationships.

### Limitations and Future work

This study had some limitations. First, there was no questionnaire on the agent's appearance, which is a limitation in exploring the potential effect of an agent's human likeness. Including the Godspeed questionnaire<sup>71</sup> as a potentially more sensitive



tool, our next step of the research will further examine the potential effect of the agent's human likeness using well-known human-agent interaction questionnaires. Second, many more males than females were recruited in both experiments. Although gender is not a factor that significantly affects the results of this study, it would be better to effectively control this factor. We will enroll a larger sample and well control the individual factors in future work. Third, this study used the CAVE system<sup>38</sup> to create an immersive VR experimental environment. Although wearing a VR head-mounted display (HMD)<sup>72</sup> can also create an immersive environment and is more accessible to researchers, two considerations halt this study from choosing it: 1) participants need to wear an EEG headband, which may interfere with VR HMD while wearing; 2) the experimental time in this study is slightly long (12 min/round) for VR HMD, which may lead to visual fatigue or VR sickness, but this is not a serious problem for the CAVE system<sup>72</sup>. We believe that if these two considerations can be further optimized, our designed task can be expanded to VR HMD, which will be more available and beneficial to other researchers in future studies.

## Conclusions

In this study, we explore the effect of agents' social cues on the human-agent relationship and performance in the tedious task of parcel sorting. Two social cues, that is, human likeness and responsiveness of the agent, were examined by constructing a virtual parcel-sorting task and a prototype system. Our results show that:

- The responsiveness of the agent has a significant effect on the user experience in a tedious task. Responsiveness and an appropriate social response strategy to show responsiveness can greatly improve the human-agent relationship without compromising the performance efficiency of tedious tasks.
- Compared with human likeness, the responsiveness of the agent is more important for building a human-agent relationship.

In summary, agents that have a higher level of responsiveness and use appropriate response strategies to show responsiveness to users may be most conducive to building good human-agent relationships in tedious tasks.

## Data availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

## References

1. Franklin, S. & Graesser, A. Is it an agent, or just a program?: A taxonomy for autonomous agents. In *International workshop on agent theories, architectures, and languages*, 21–35 (Springer, 1996).
2. Oertel, C. *et al.* Engagement in human-agent interaction: An overview. *Front. Robotics AI* **7**, 92 (2020).
3. Emmerich, K., Ring, P. & Masuch, M. I'm glad you are on my side: How to design compelling game companions. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 141–152 (2018).
4. Rato, D. & Prada, R. A taxonomy of social roles for agents in games. In *International Conference on Entertainment Computing*, 75–87 (Springer, 2021).
5. Abdulrahman, A., Richards, D., Ranjartabar, H. & Mascarenhas, S. Verbal empathy and explanation to encourage behaviour change intention. *J. on Multimodal User Interfaces* **15**, 189–199 (2021).
6. Hatami, J., Sharifian, M., Noorollahi, Z. & Fathipour, A. The effect of gender, religiosity and personality on the interpersonal distance preference: a virtual reality study. *Commun. Res. Reports* **37**, 182–192 (2020).
7. Shiban, Y. *et al.* The appearance effect: Influences of virtual agent features on performance and motivation. *Comput. Hum. Behav.* **49**, 5–11 (2015).
8. Bian, Y. *et al.* Effects of pedagogical agent's personality and emotional feedback strategy on chinese students' learning experiences and performance: a study based on virtual tai chi training studio. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 433–444 (2016).
9. Lopes, D. S., dos Anjos, R. K. & Jorge, J. A. Assessing the usability of tile-based interfaces to visually navigate 3-d parameter domains. *Int. J. Human-Computer Stud.* **118**, 1–13 (2018).
10. Rackley, E. D. & Kwok, M. "long, boring, and tedious": youths' experiences with complex, religious texts. *Literacy* **50**, 55–61 (2016).
11. Li, X. *et al.* Mobile phone-based device for personalised tutorials of 3d printer assembly. In *International Conference on Human-Computer Interaction*, 37–48 (Springer, 2019).

12. Mumm, J. & Mutlu, B. Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Comput. Hum. Behav.* **27**, 1643–1650 (2011).
13. Lester, J. C. *et al.* The persona effect: affective impact of animated pedagogical agents. In Pemberton, S. (ed.) *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 359–366 (1997).
14. Tien, L. T. & Osman, K. Pedagogical agents in interactive multimedia modules: Issues of variability. *Procedia-Social Behav. Sci.* **7**, 605–612 (2010).
15. Kim, Y., Baylor, A. L. & Shen, E. Pedagogical agents as learning companions: the impact of agent emotion and gender. *J. Comput. Assist. Learn.* **23**, 220–234 (2007).
16. Ozogul, G., Johnson, A. M., Atkinson, R. K. & Reisslein, M. Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Comput. & Educ.* **67**, 36–50 (2013).
17. van der Meij, H., van der Meij, J. & Harmsen, R. Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. *Educ. technology research development* **63**, 381–403 (2015).
18. Wiese, E. & Weis, P. P. It matters to me if you are human-examining categorical perception in human and nonhuman agents. *Int. J. Human-Computer Stud.* **133**, 1–12 (2020).
19. Krach, S. *et al.* Can machines think? interaction and perspective taking with robots investigated via fmri. *PloS one* **3**, e2597 (2008).
20. Castelfranchi, C. Modelling social action for ai agents. *Artif. intelligence* **103**, 157–182 (1998).
21. Hofstede, G. J., Frantz, C., Hoey, J., Scholz, G. & Schröder, T. Artificial sociality manifesto. *Rev. Artif. Soc. Soc. Simulation*, 8th April (2021).
22. Liew, T. W. & Tan, S.-M. Virtual agents with personality: Adaptation of learner-agent personality in a virtual learning environment. In *2016 Eleventh International Conference on Digital Information Management (ICDIM)*, 157–162 (IEEE, 2016).
23. Beale, R. & Creed, C. Affective interaction: How emotional agents affect users. *Int. journal human-computer studies* **67**, 755–776 (2009).
24. Bickmore, T. W. & Picard, R. W. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Comput. Interact. (TOCHI)* **12**, 293–327 (2005).
25. Cafaro, A., Vilhjálmsson, H. H. & Bickmore, T. First impressions in human-agent virtual encounters. *ACM Transactions on Comput. Interact. (TOCHI)* **23**, 1–40 (2016).
26. Zhao, G., Liu, Y.-J. & Shi, Y. Real-time assessment of cross-task mental workload using physiological measures during anomaly detection. *IEEE Transactions on Human-Machine Syst.* **48**, 149–160 (2018).
27. Lazar, J., Feng, J. H. & Hochheiser, H. *Research methods in human-computer interaction* (Morgan Kaufmann, 2017).
28. Ay, B. *et al.* Automated depression detection using deep representation and sequence learning with eeg signals. *J. medical systems* **43**, 1–12 (2019).
29. Meiselwitz, G. *Social Computing and Social Media. Communication and Social Communities: 11th International Conference, SCSM 2019, Held as Part of the 21st HCI International Conference, HCHI 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part II*, vol. 11579 (Springer, 2019).
30. Hoppe, M. *et al.* A human touch: Social touch increases the perceived human-likeness of agents in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–11 (2020).
31. Kim, Y. & Wei, Q. The impact of learner attributes and learner choice in an agent-based environment. *Comput. & Educ.* **56**, 505–514 (2011).
32. Reeves, B. & Nass, C. *The media equation: How people treat computers, television, and new media like real people* (Cambridge university press Cambridge, United Kingdom, 1996).
33. Moon, Y. & Nass, C. How “real” are computer personalities? psychological responses to personality types in human-computer interaction. *Commun. research* **23**, 651–674 (1996).
34. Kane, H. S., McCall, C., Collins, N. L. & Blascovich, J. Mere presence is not enough: Responsive support in a virtual world. *J. experimental social psychology* **48**, 37–44 (2012).
35. Złotowski, J., Strasser, E. & Bartneck, C. Dimensions of anthropomorphism: from humanness to humanlikeness. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 66–73 (IEEE, 2014).

36. Ruijten, P. A., Haans, A., Ham, J. & Midden, C. J. Perceived human-likeness of social robots: testing the rasch model as a method for measuring anthropomorphism. *Int. J. Soc. Robotics* **11**, 477–494 (2019).
37. Nilakantan, J. M., Huang, G. Q. & Ponnambalam, S. G. An investigation on minimizing cycle time and total energy consumption in robotic assembly line systems. *J. Clean. Prod.* **90**, 311–325 (2015).
38. Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V. & Hart, J. C. The cave: audio visual experience automatic virtual environment. *Commun. ACM* **35**, 64–73 (1992).
39. Denisova, A., Cairns, P., Guckelsberger, C. & Zendle, D. Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (corgis). *Int. J. Human-Computer Stud.* **137**, 102383 (2020).
40. Patel, K., Shah, H., Dcosta, M. & Shastri, D. Evaluating neurosky's single-channel eeg sensor for drowsiness detection. In Stephanidis, C. (ed.) *HCI International 2017 – Posters' Extended Abstracts*, 243–250 (Springer International Publishing, 2017).
41. G, J. *et al.* Automated detection of absence seizures using a wearable electroencephalographic device: a phase 3 validation study and feasibility of automated behavioral testing. *Epilepsia* **00**, 1–7 (2022).
42. Navalyal, G. U. & Gavvas, R. D. A dynamic attention assessment and enhancement tool using computer graphics. *Human-centric Comput. Inf. Sci.* **4**, 1–7 (2014).
43. Wu, S.-F., Lu, Y.-L. & Lien, C.-J. Detecting students' flow states and their construct through electroencephalogram: Reflective flow experiences, balance of challenge and skill, and sense of control. *J. Educ. Comput. Res.* **58**, 1515–1540 (2021).
44. Bangor, A., Kortum, P. T. & Miller, J. T. An empirical evaluation of the system usability scale. *Intl. J. Human-Computer Interact.* **24**, 574–594 (2008).
45. Jennett, C. *et al.* Measuring and defining the experience of immersion in games. *Int. journal human-computer studies* **66**, 641–661 (2008).
46. Cairns, P., Cox, A. & Nordin, A. I. Immersion in digital games: review of gaming experience research. *Handb. digital games* 337–361 (2014).
47. McQuiggan, S. W., Robison, J. L., Phillips, R. & Lester, J. C. Modeling parallel and reactive empathy in virtual agents: an inductive approach. In *AAMAS (1)*, 167–174 (Citeseer, 2008).
48. Hoffman, M. L. *Empathy and moral development: Implications for caring and justice* (Cambridge University Press, 2001).
49. Li, J. *et al.* Study on horse-rider interaction based on body sensor network in competitive equestrian. *IEEE Transactions on Affect. Comput.* (2019).
50. NeuroSky Inc. *NeuroSky's eSense™ meters and detection of mental state.* (2009). Technical report.
51. Chen, H.-M., Chen, S.-Y., Jheng, T.-J. & Chang, S.-C. Design of a mobile brain-computer interface system with personalized emotional feedback. In *Future Information Technology-II*, 87–95 (Springer, 2015).
52. Liu, C. C. A model for exploring players flow experience in online games. *Inf. Technol. & People* **30**, 139–162 (2017).
53. Novak, T. P., Hoffman, D. L. & Yung, Y.-F. Measuring the customer experience in online environments: A structural modeling approach. *Mark. science* **19**, 22–42 (2000).
54. Flores Jr, L. E. & Berenbaum, H. Desire for emotional closeness moderates the effectiveness of the social regulation of emotion. *Pers. individual differences* **53**, 952–957 (2012).
55. Wang, J., Yang, H., Shao, R., Abdullah, S. & Sundar, S. S. Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–13 (Association for Computing Machinery, New York, NY, USA, 2020).
56. Weisman, O., Aderka, I. M., Marom, S., Hermesh, H. & Gilboa-Schechtman, E. Social rank and affiliation in social anxiety disorder. *Behav. Res. Ther.* **49**, 399–405 (2011).
57. Flores, L. E. & Berenbaum, H. Desired emotional closeness moderates the prospective relations between levels of perceived emotional closeness and psychological distress. *J. Soc. Clin. Psychol.* **33**, 673–700 (2014).
58. Chen, J. & Zhou, X. Within-family patterns of intergenerational emotional closeness and psychological well-being of older parents in china. *Ageing & Mental Heal.* **25**, 711–719 (2021).

59. Wu, P.-C., Foo, M.-D. & Turban, D. B. The role of personality in relationship closeness, developer assistance, and career success. *J. Vocat. Behav.* **73**, 440–448 (2008).
60. Levin, I. P., Proskansky, C. M., Heller, D. & Brunick, B. M. Prescreening of choice options in ‘positive’ and ‘negative’ decision-making tasks. *J. Behav. Decis. Mak.* **14**, 279–293 (2001).
61. Xanthopoulos, P. & Razzaghi, T. A weighted support vector machine method for control chart pattern recognition. *Comput. & Ind. Eng.* **70**, 134–149 (2014).
62. Galla, B. M. *et al.* The academic diligence task (adt): Assessing individual differences in effort on tedious but important schoolwork. *Contemp. educational psychology* **39**, 314–325 (2014).
63. Wickens, C. D. Processing resources and attention. In *Multiple-task performance*, 3–34 (CRC Press, 2020).
64. Potdevin, D., Clavel, C. & Sabouret, N. Virtual intimacy in human-embodied conversational agent interactions: the influence of multimodality on its perception. *J. on Multimodal User Interfaces* **15**, 25–43 (2021).
65. Cassell, J. & Bickmore, T. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Model. User-Adapted Interact.* **13**, 89–132 (2003).
66. De Graaf, M. M. & Allouch, S. B. Exploring influencing variables for the acceptance of social robots. *Robotics autonomous systems* **61**, 1476–1486 (2013).
67. Huang, L., Morency, L. & Gratch, J. Virtual rapport 2.0. In *Springer Berlin Heidelberg*, 68–79 (2011).
68. Zhao, R., Romero, O. J. & Rudnicky, A. Sogo: A social intelligent negotiation dialogue system. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents, IVA '18*, 239–246 (Association for Computing Machinery, New York, NY, USA, 2018).
69. Gratch, J. *et al.* Virtual rapport. In *International Workshop on Intelligent Virtual Agents*, 14–27 (Springer, 2006).
70. Kiesler, S. & Goetz, J. Mental models and cooperation with robotic assistants. retrieved on november 24, 2004 (2002).
71. Bartneck, C., Kulić, D., Croft, E. & Zoghbi, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. journal social robotics* **1**, 71–81 (2009).
72. Sharples, S., Cobb, S., Moody, A. & Wilson, J. R. Virtual reality induced symptoms and effects (vrise): Comparison of head mounted display (hmd), desktop and projection display systems. *Displays* **29**, 58–69 (2008).

## Author contributions statement

Chao Zhou, Yulong Bian and Yong-Jin Liu conceived the experiment(s), Ziyang Zhang and Yaoyuan Wang prepared the hardware and software, Chao Zhou and Yulong Bian conducted the experiment(s), Chao Zhou and Yulong Bian analysed the results, Chao Zhou and Yulong Bian wrote the original draft, Yong-Jin Liu and Shu Zhang revised the draft. All authors reviewed the manuscript.

## Acknowledgment

This work was partially supported by the postdoctoral research foundation of china (2021TQ0178), Tsinghua University Initiative Scientific Research Program and the Natural Science Foundation of China (61725204).

## Additional information

**Competing interests** No competing interests exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part.