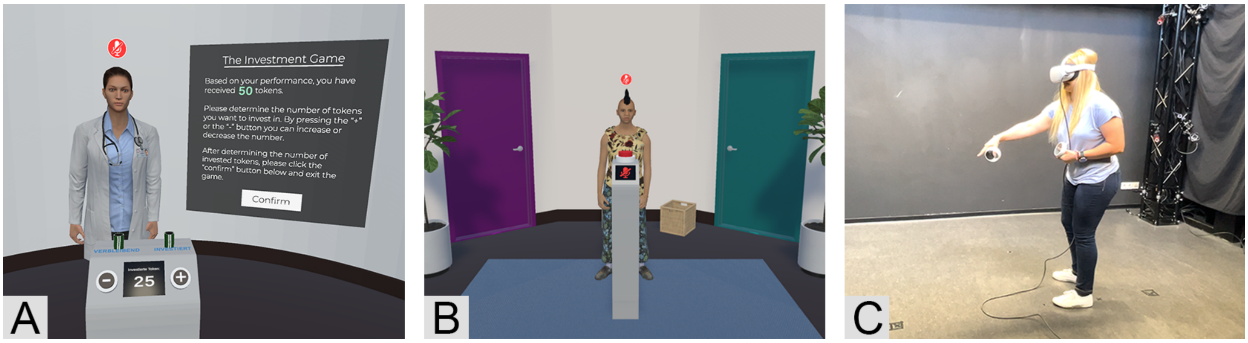
Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm

Jinghuai Lin*, Johrine Cronjé*, Ivo Käthner, Paul Pauli, and Marc Erich Latoschik


The Investment Game

Based on your performance, you have received 50 tokens.

Please determine the number of tokens you want to invest in. By pressing the "+" or the "-" button you can increase or decrease the number.

After determining the number of invested tokens, please click the "confirm" button below and exit the game.

Fig. 1: We used the combination of the investment game and the virtual maze paradigm to measure users' trust towards virtual humans. A) In the investment game, participants need to decide the number of tokens they want to invest in the trustee. B) In the virtual maze, participants must select one of the two doors to escape when entering a new room. The virtual human appears in the middle of the room. C) A participant interacts with the virtual environment in the virtual maze in an immersive VR experience.

Abstract—Virtual humans, including virtual agents and avatars, play an increasingly important role as VR technology advances. For example, virtual humans are used as digital bodies of users in social VR or as interfaces for AI assistants in online financing. Interpersonal trust is an essential prerequisite in real-life interactions, as well as in the virtual world. However, to date, there are no established interpersonal trust measurement tools specifically for virtual humans in virtual reality. This study fills the gap, by contributing a novel validated behavioural tool to measure interpersonal trust towards a specific virtual social interaction partner in social VR. This validated paradigm is inspired by a previously proposed virtual maze task that measures trust towards virtual characters. In the current study, a variant of this paradigm was implemented. The task of the users (the trustors) is to navigate through a maze in virtual reality, where they can interact with a virtual human (the trustee). They can choose to 1) ask for advice and 2) follow the advice from the virtual human if they want to. These measures served as behavioural measures of trust. We conducted a validation study with 70 participants in a between-subject design. The two conditions did not differ in the content of the advice but in the appearance, tone of voice and engagement of the trustees (alleged as avatars controlled by other participants). Results indicate that the experimental manipulation was successful, as participants rated the virtual human as more trustworthy in the trustworthy condition than in the untrustworthy condition. Importantly, this manipulation affected the trust behaviour of our participants, who, in the trustworthy condition, asked for advice more often and followed advice more often, indicating that the paradigm is sensitive to assessing interpersonal trust towards virtual humans. Thus, our paradigm can be used to measure differences in interpersonal trust towards virtual humans and may serve as a valuable research tool to study trust in virtual reality.

Index Terms—virtual human, specific interpersonal trust, trustworthiness, social VR, behavioural measurement paradigm, virtual reality

1 Introduction

With the advance of VR technology, the growing public interest and the huge investments by tech giants in recent years, virtual reality has gradually entered different aspects of our lives, including entertainment [61, 75], social interaction [63], education [18, 58], healthcare [25, 68], the workplace [42] and many more. In the development of these VR applications, virtual humans play an important role. Virtual humans can either be controlled by humans, as avatars that represent users; or be controlled by algorithms, often in the form of intelligent virtual agents (IVAs). Virtual humans have various applications: for instance, in social VR [45], virtual humans can be utilized as the digital bodies of users, allowing users to be immersed in cyberspace to communicate, interact, and collaborate with each other [63]. In many scenarios, such as online healthcare or financing, virtual humans can act as AI assistants, enhancing their social presence [40] and potentially increasing users' acceptance. The development of virtual humans, especially those with a realistic appearance, serves the purpose of enhancing virtual embodiment [62], recreating the real world [71], and making the virtual world an alternative realm for human socio-cultural activities [15].

- ** Jinghuai Lin and Johrine Cronjé contributed equally to the work.*
- *Jinghuai Lin is with the Human-Computer Interaction (HCI) Group at the University of Würzburg. E-mail: jinghuai.lin@uni-wuerzburg.de.*
- *Johrine Cronjé is with the Department of Psychology I, Biological Psychology, Clinical Psychology and Psychotherapy at the University of Würzburg. E-mail: johrine.cronje@uni-wuerzburg.de.*
- *Ivo Käthner is with the Department of Psychology I, Biological Psychology, Clinical Psychology and Psychotherapy at the University of Würzburg, and the Department of Physiological Psychology at the University of Bamberg. E-mail: ivo.kaethner@uni-wuerzburg.de.*
- *Paul Pauli is with the Department of Psychology I, Biological Psychology, Clinical Psychology and Psychotherapy at the University of Würzburg, and the Center of Mental Health, Medical Faculty at the University of Würzburg. E-mail: pauli@psychologie.uni-wuerzburg.de.*
- *Marc Erich Latoschik is with the Human-Computer Interaction (HCI) Group at the University of Würzburg. E-mail: marc.latoschik@uni-wuerzburg.de.*

Manuscript received 14 October 2022; revised 13 January 2023; accepted 30 January 2023. Date of publication 22 February 2023; date of current version 29 March 2023. Digital Object Identifier no. 10.1109/TVCG.2023.3247095

No lasting relationship can be established and maintained without trust [48, 50], either in real life or in the virtual world. Users in social VR are vulnerable and at risk in relationships without interpersonal trust, whether it is to transact with a person or a business, or trusting another user's identity in a social environment [8, 41]. Trust has been studied from many scientific perspectives, leading to a wide range of definitions [4, 10, 21, 27, 50]. There are different categories of trust, and it consist of multiple factors, for example, interpersonal trust (generalized or specific), affect-based trust, cognitive-based trust, trust in automation, and trust in business [49]. With social interaction taking place more frequently in VR, it is essential to ask which category this "type" of trust towards virtual humans belongs to.

We consider trust towards virtual humans (both avatars and agents) as interpersonal trust. In the case of avatars, they can be considered extensions of users [19, 20, 24], with social interactions between avatars actually taking place between users. In the case of intelligent virtual agents, researchers suggest that the human-computer relationship is fundamentally social and consists of the same social norms as a human-to-human relationship does [51]. We resonate with the definition of trust by Lee and See [39]: "the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability[1]". It refers to the dyad relationship between one person and another specific social interaction partner.

The measurement of interpersonal trust between virtual social interaction partners (virtual humans) requires researchers' attention. First, the attribution of human characteristics and human behaviour to virtual humans could impact trust. Such human factors include cooperativeness [49], physical appearance (facial expressions, avatar behaviour, tone of voice) [43], anthropomorphism, and trust resilience [14]. Measuring trust helps to investigate the interplay between such factors in trust building. Second, the measurement of trust is essential when comparing human-to-human interaction with human-to-virtual human interaction, which could result in a better understanding of social responses to virtual human interfaces and improve interface design [78]. For example, it has been found that using self-avatars in shared virtual environments can lead to an increase of trust formation in collaborations, compared to using only the model of VR controllers as representation [53]. Lastly, compared to people in real life, virtual humans can be modified in appearance and voice, potentially resulting in changes in trust evaluations. For instance, identity theft has been a concern in social VR [41], where cybercriminals can steal and control other users' avatars to gain trust and profits; virtual agents can intentionally be designed to appear more trustworthy to convince customers for commercial reasons. Such variability makes the study of trust towards virtual humans a priority.

Several measures to assess generalized trust (how one trusts others in general) have been previously proposed. However, there is a lack of instruments for measuring specific trust (trust towards a specific person in a specific circumstance). This holds particularly true for paradigms suitable for measuring interpersonal trust towards virtual humans in virtual reality. Hale et al. [26] proposed a virtual maze task that relies on advice-seeking behaviour to measure trust towards two opposing characters. However, in the proposed form, their paradigm is unsuitable for measuring interpersonal trust towards virtual humans as social interaction partners due to social desirability bias. Additionally, their implementation could have been enhanced by utilizing the full capabilities of modern VR experiences and incorporating fundamental VR characteristics such as immersion, virtual embodiment, and presence. This may have improved the generalizability of their results to social VR contexts.

The presented work addressed these points. The motivation for an improved version has the VR community in mind and is rooted in the need for an in-the-moment interpersonal trust measurement tool that takes fundamental inherent VR characteristics into account. This tool should be sensitive to the manipulation of the virtual humans' trustworthiness that ultimately guides trusting behaviour. Our contribution includes a paradigm for investigating interpersonal trust towards virtual agents and avatars in virtual reality. In comparison to Hale et al.'s paradigm [26], our virtual humans have more realistic social cues, and our task emphasizes the immersion and interactivity of VR. Our paradigm includes one specific virtual human during the maze task. Participants' behavioural differences stem from their subjective evaluation of the virtual human, instead of comparing two virtual humans, which might lead to social desirability bias instead of trust evaluation. Furthermore, our paradigm includes a believable and self-contained cover story that keeps participants engaged through a motivational incentive in the midst of uncertainty. With a validation study, we concluded that our paradigm is sensitive to the manipulation of trustworthiness and can be used to measure differences in interpersonal trust towards virtual humans.

2 Related Work

2.1 Trust measurements

Measurements of trust in experimental research are generally devided into two categories: subjective and objective measures.

Subjective measurements of trust are primarily self-report questionnaires, the predominant method to measure trust across different domains in psychology, neuroscience, sociology, and organization science [26]. These include the "Interpersonal Trust scale" (ITS) developed by Rotter [64], and other alternatives such as the "General Trust Scale" (GTS) by Yamagishi and Yamagishi [77], the "KUSIV3" by Beierlein et al. [6], to name a few. In addition, other relevant scales, including the "Self-disclosure Index" (SDI) [47], the "European Social Survey" (ESS) [57], and others, are often used as additional measures of trust. However, most of these questionnaires only measure generalized trust [13]—a reflection of how much a person trusts others in general [26], rather than specific trust—trust towards a specific person, either to people with close relationships or strangers [26]. Robbins [59], on the other hand, constructed the "Stranger-Face Trust" (SFT) questionnaire that aims to measure trust in specific strangers and particular matters.

As Chan [11] has pointed out, most self-report methods reflect internal feelings less accurately. Subjective trust measurements are considered poor predictors of external behaviour [2, 22, 44], and may not be ideal for measuring specific trust as there can be multiple interpretations of items and trust [7]; thus, objective measurements are preferred. Objective methods often use behavioural clues during social interactions as proxies of trust. The Trust Game [36] is one of the most popular and established measures of trust in behavioural economics and psychological research. In the Trust Game, the trustor can transfer a certain fraction p of a monetary endowment given to the trustee, while the transferred fraction is magnified by a factor $K > 1$ (e.g., doubled or tripled) before sending it to the trustee. The trustee can then return a certain fraction q of the received amount to the trustor. However, there is no guarantee of such a return. In this paradigm, trust is measured by the fractions of transfers during the back and forth, with the expectation of a significant sum in return while risking the possibility that no reward will be returned. Similar ideas are adopted by a variant of the Trust Game or similar paradigms, including the Dictator Game [29], and the Investment Game [9]. Additionally, research indicates that interpersonal distance between social interaction partners, advice-seeking behaviour, and the duration of mutual gaze [3, 12, 26, 55, 60] can be indicators of trust.

2.2 Measuring trust towards virtual humans

Both subjective and objective measurements have been used for measuring trust towards virtual humans such as avatars or agents. Most research on trust towards avatars relied on self-reports as the primary measurement [40, 70] or combined self-reports with other

[1] In this definition, an agent can be automation or another person that actively interacts with the environment on behalf of the person [39].

measures [3,26,52]. As for objective measurements, Bente et al. [8] investigated how photorealistic avatars and reputation scores affect trust-building in online transactions using the Trust Game. Hale et al. [26] implemented the Investment Game to test specific trust towards interactive virtual characters. However, they found that the results of different characters are highly correlated, which suggests that the Investment Game measures generalized trust rather than specific trust [26].

As alternatives, advice-seeking behaviour and the ask–endorse paradigm [12,26,52] have recently been examined as new approaches to measure trust. Such methods measure whether participants will seek and follow advice or information from a specific person. For example, Pan and Steed [52] conducted a comparison study of trust among avatar-, video-, and robot-mediated interaction by asking participants to complete a quiz, and recording the number of times they asked for and followed advice from two advisors randomly selected from the three alternative representations. Hale et al. [26] also adopted the ask-endorse paradigm to measure specific trust towards virtual characters using a Virtual Maze task. Their work has inspired the design of our paradigm and will be further explained in the subsection below.

2.3 The *Virtual Maze*

Focusing on the measurement of generalized trust versus specific trust, Hale et al. [26] proposed a novel behavioural task, "the virtual maze", inspired by the ask-endorse paradigm to measure trust between users and virtual agents through behavioural proxies of trust [32–35, 54].

In the virtual maze task, participants must navigate through a virtual maze of identical rooms. When entering a "new room", they are told to select one of the two doors in front of them to escape. To assist them in their decision-making, two virtual characters are present to provide navigation advice if the participants decide to approach them (optional). When the virtual characters are approached, they will suggest a door. The participants keep making decisions until they are notified that they have escaped from the maze. Unknown to the participants, there are no right or wrong decisions on the way out of the maze. Rooms and corridors are automatically generated until enough trials (rooms) are observed, and the participant has supposedly escaped. The trust towards each character is measured by 1) the number of times each virtual character is approached for advice and 2) the number of times participants followed the advice of each character.

The trustworthiness of virtual characters was manipulated through brief interviews in which the participants asked the characters prepared questions before the maze task. As a result, their verbal answers and non-verbal vocal behaviours differ so that one character appears trustworthy and the other appears untrustworthy. They have also included subjective ratings as validation measures and compared them with behavioural measures in an investment game. Their results indicate that participants followed advice from the trustworthy character significantly more than the untrustworthy character. Furthermore, trust behaviour in the virtual maze task shows no correlation between the two characters, indicating that it only reflects specific trust. In comparison, behaviour in the investment game reflected both specific trust and generalized trust.

The virtual maze allows the measurement of trust towards a specific virtual character rather than the propensity to trust others in general. Thus, it could help measure interpersonal trust towards virtual humans. However, certain adaptations are needed. First, virtual humans include virtual agents and avatars which are usually considered the proxies of AI or humans. Considering the definition of interpersonal trust, virtual humans as trustees are expected to have an independent will and intelligence to give their answers, which creates uncertainty and vulnerability during interactions. In the design of Hale et al. [26], participants perceived the trustees as pre-scripted characters that reflect the preliminary design of the experiment and the will of the experimenters. This could potentially lead to different trust constructs: the trust behaviour being measured could stem from whether they believe the backgrounds and personalities assigned to the characters, their interpretations of the experiments, or their trust towards the experimenter. For example, participants could assume that the experimenters are controlling the advice given. Second, in the design of Hale et al. [26], two distinguished characters always appeared together in each room and provided opposing advice. The decision to follow the advice as a proxy of trust largely relies on the comparison between the two characters, which is impractical when we only have one virtual human. For example, the virtual maze cannot be used when measuring interpersonal trust towards one specific avatar or investigating which external factors influence trust behaviours. In addition, the study design was a within-subject design. Hence, the study's goal might have been obvious to the participants and a social desirability bias might have driven their behaviour.

Although the authors claimed that the task was designed for virtual reality, they did not utilize the full capabilities of immersion and interactivity of modern VR experiences. In their three studies, the virtual environment and characters were displayed with a projector, an head-mounted display (HMD), and a desktop PC, respectively. In either version, participants can only navigate with a joystick to approach the virtual characters. Compared to realistic locomotion and interaction in social VR nowadays, where participants can "walk around", move and use their "virtual hands" to interact with the environment, their relatively low level of immersion and agency could hinder the virtual embodiment, body ownership and presence [28,62]. Given the often reported effects of virtual embodiments and presence on secondary factors such as emotional response [74] and especially trust [65], we argue that their results cannot be easily generalized to modern VR experiences. Thus, a paradigm that better reflects modern VR is desired.

Additionally, we noticed some limitations that need to be improved in the design of Hale et al. [26]: 1) The rooms are identical; participants may instead feel that they are staying in the same place when entering new rooms. 2) It is unclear to the participants whether it is still possible to escape if they have previously made a wrong decision. Such uncertainty may result in a loss of motivation. 3) These virtual characters lack realistic social cues, such as eye contact, which can lead to low congruency, plausibility, and social presence.

2.4 Summary

To summarize, the previous work provides several measurements of trust, including subjective measurements, such as self-report questionnaires and objective measurements that use behavioural paradigms such as the Trust Game. However, most trust measurements are only suitable for measuring generalized trust rather than specific trust. Among them, the work of Hale et al. [26] has inspired our research. They have created a novel virtual maze task to behaviourally measure trust towards virtual characters. For the current study, we utilize their design while providing improvements with the following goals to address the limitations of their work: 1) to have a suitable design for the VR community with consideration of fundamental VR characteristics such as immersion, embodiment, and presence; 2) to investigate interpersonal trust towards agents and avatars as social interaction partners rather than being limited to pre-scripted agents and behaviour driven by social desirability bias; 3) to be more adaptive for future investigation of trust; and 4) to have a believable and self-contained cover story that could provide a strong framing effect and motivational incentives during the maze task.

3 Design and Hypotheses

Building on the work of Hale et al. [26] and considering their limitations, we implemented an improved version of the virtual maze paradigm, aiming to measure trust with the behavioural proxies: 1) how often the trustor asks, 2) follows the advice, and 3) the time spent before they made their decision. A validation study was conducted to verify whether our paradigm can measure differences in interpersonal trust towards virtual humans.

3.1 The virtual maze paradigm

The motivation for creating such a paradigm is to study the trust relationships between users in VR applications, especially in social VR. It is crucial to provide an immersive environment with a high level of immersion and interactivity; virtual humans (either controlled by real people or pre-scripted) need to be regarded as social interaction partners with a high level of realism and social presence. Thus, our version of the virtual maze is implemented as an immersive VR experience. Participants wear an HMD and interact with the virtual humans and the virtual environment with the VR controllers. Although the design and setups described below are primarily for measuring trust towards (alleged) avatars, the proposed paradigm can be adapted for agents. Implications for such adaptations are mentioned where applicable.

The introductory video. As a first step, participants will watch an introductory video, in which the experimental procedure and the background information of social interaction through virtual humans in social VR will be explained. Participants (the trustors) will be informed that there is another participant (the trustee) who will later control an avatar and join them in the maze. The video explains how the trustee will participate in the experiment. Supposedly the trustee will talk through a microphone to provide advice if the trustors unmute them by pressing a virtual button. In reality, the trustee can either be a real person behind the avatar or a pre-scripted agent. Nevertheless, participants need to believe that they will later be interacting with a real person behind the virtual human. As we have mentioned, this is crucial for participants' decisions and trust behaviour in the maze to be driven by the interpersonal trust towards the virtual human. In addition, an alleged map of the maze will be shown to participants, informing them that there are many escape routes, that they need to escape as quickly as possible, and that the trustee could refer to the map to advise them. Importantly, participants are explicitly informed that the trustee is given the freedom to mislead or lie to the trustors if they want to and that it is of no advantage to them to help or mislead the trustors (to create uncertainty and ambiguity).

The introductory video is an essential component of our paradigm. It serves three purposes: 1) to foster the idea of having a real person behind the virtual human so that participants perceive the trustee as an interaction partner; 2) to inform participants that there are multiple escape routes and that the trustee can provide advice according to where they are currently located; 3) to emphasize that the trustee has a choice to either be helpful or misleading to prevent participants from blindly following all the given advice.

When using this paradigm to measure trust towards an intelligent agent, the video can be adapted by showing an introduction of the agent and how it is able to communicate and provide advice.

VR tutorial and the testing room. After watching the introductory video, participants will put on the HMD and VR controllers and enter the tutorial task. In the tutorial participants practice how to use the VR controllers to open the doors, navigate through the maze, and unmute the trustee to get advice.

After the tutorial, participants will enter the testing room where a connection with the trustee will be established, and they will see the avatar for the first time. In the testing room, participants are asked to unmute the trustee so that the trustee can talk to them. If the trustee is an avatar with a real person behind or an intelligent agent, the testing room can be used to ensure that the connection has been established. When a pre-scripted agent is implemented, the pre-recorded speech of the agent is played to pretend that the connection has been established.

The virtual maze task. After the testing room, participants will start the virtual maze task. The maze consists of several rooms, and in each room, there are two doors opposite the participants which lead to different rooms. As illustrated in Figure 1B, the trustee avatar will appear in the middle of the room, as it does not imply preference of which door to choose. Participants can walk around freely in the room. To open a door, participants need to point to the door with a controller and press a button (Figure 3A). When they decide to ask for advice, they need to approach the avatar and press the red button in front of it with their virtual hands to unmute them (Figure 2 and Figure 3B). It is simpler and quicker in operation to directly open a door than to ask the avatar for advice, by design, so it is less likely that participants will blindly seek for advice all the time. When the trustee is unmuted, the status symbol will turn green, and the trustee can talk to the participants to give advice. In the case of a pre-scripted agent, a pre-recorded audio will be played. We also inform participants that the trustee cannot hear them to avoid attempts at conversations.

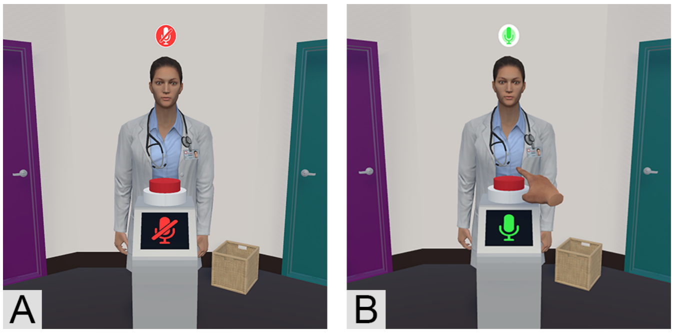


Fig. 2: A) Every time participants enter a new room, the virtual human will be muted. B) Participants can press the red button in front of the virtual human to unmute them if they ask for advice.

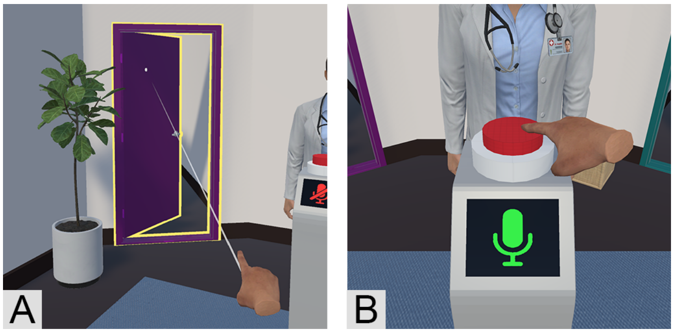


Fig. 3: A)Participants can point to a door and open it by pressing a button on their controllers. B)To seek advice from the virtual human, participants need to approach and unmute them.

Once participants select a door, they will be teleported to a new room. To give participants the impression that they are entering a new room at each trial, minor changes to the virtual environment were made (for example, changing the colour of the doors, adding and/or removing plants or objects, etc.). The trustee avatar will appear as muted again in the middle of the room. Unknown to participants, there is no right or wrong door to choose from. After participants decide and enter a new room 12 times (12 trials), they will be informed that they have escaped the maze. Since participants receive no feedback on their progress in the maze, their experience of uncertainty is increased, and their decision-making is uninfluenced by external motivators.

In each room, we record 1) whether they ask for advice; 2) whether they follow the advice given (if applicable); and 3) the response time of following the advice (if applicable).

Table 1: Manipulation of the virtual humans' trustworthiness in the validation study.

	Trustworthy	Untrustworthy
Appearance	high trust rating	low trust rating
Voice & Tone	patient, motivated	impatient, unmotivated
Engagement	already waiting for the participants; focus on the task	30 seconds late; try to talk to the experimenter

3.2 The validation study

3.2.1 Manipulation of trustworthiness

To investigate whether our paradigm can measure differences between interpersonal trust towards virtual humans, we designed a between-subject study for validation. First, we manipulated the trustworthiness of the avatars in the two conditions. The manipulations included the avatars' appearance, tone of voice, and engagement during the task (Table 1). To ensure that all participants are exposed to the same contents and interactions, we used pre-scripted agents instead of having a real person control the avatars. However, participants were under the impression that they were interacting with real humans.

The avatars' appearance was selected through a pre-study, where several avatars were rated. The avatar with the highest trustworthiness score was chosen for the trustworthy condition, and the avatar with the lowest trustworthiness score was chosen for the untrustworthy condition. Both conditions did not differ in the content of the advice but varied in the avatars' tone of voice. For example, the voice of the trustworthy agent gave the impression of someone who is patient, motivated, gentle and tender tempered, whereas the voice of the untrustworthy agent gave the opposite impression of someone who is impatient, unmotivated, abrupt and unaffected. Additionally, in the trustworthy condition, the avatar was already waiting for participants in the testing room; they greeted them politely and gave undivided attention during the experiment; on the contrary, in the untrustworthy condition, the avatar was 30 seconds late to the testing session, they were distracted during the experiment and tried to talk with the experimenter in the background. Regarding these manipulations, we formulated the following research question:

RQ1: Will our manipulations of the avatars lead to differences in their perceived trustworthiness?

In addition to using the introductory video, we designed the following details to create the illusion that a real person is behind the trustee avatar to make it more believable: 1) the pre-recorded audio was made to sound natural and realistic and included hesitant pauses and filler words such as "uhm..."; 2) in the testing room, when participants unmuted the trustee, they could hear the experimenter on the other side instructing the trustee to speak to the participant; 3) during the maze task, a glitch appear (the avatar switches into a T-pose for 2 seconds) with the trustee apologizing for the glitch. We included self-report questions asking whether participants felt they were interacting with a real person, and the social presence questionnaire [5] after the tasks to evaluate the effectiveness of the framing.

3.2.2 Measurement of trust

The maze task assesses whether the following behaviours are sensitive to the differences in interpersonal trust or not: 1) how often participants ask for advice, 2) how often participants follow advice, and 3) the response time of following the advice. Thus, we formulated our second research question:

RQ2: Which behaviour(s) in the maze can be used as the measurement of interpersonal trust?

In addition to the trust behaviour in the virtual maze, we also used subjective measures (self-reports through established questionnaires, as well as self-constructed explicit ratings of trustworthiness) to check whether our manipulation was successful. Furthermore, a variation of the Trust Game, a one-shot investment game before and after the maze task as an additional behavioural measurement of trust is included, to further investigate the following questions:

RQ3: Can the investment game measure differences in interpersonal trust?

The design and purpose of the validation study could then be concluded with the question:

RQ4: Is our paradigm suitable for measuring differences in interpersonal trust towards virtual humans?

3.2.3 Hypotheses

H1: The manipulation of trustworthiness is successful; participants will rate the avatar as more trustworthy in the trustworthy condition compared with the untrustworthy condition.

If **H1** proves to be true, we propose the following hypotheses for the behavioural measures in the maze task:
H2: Participants in the trustworthy condition will ask for advice significantly more often than those in the untrustworthy condition.
H3: Participants in the trustworthy condition will follow advice significantly more often than those in the untrustworthy condition.
H4: Participants in the trustworthy condition will respond quicker to advice than those in the untrustworthy condition.

For the investment game, we expected the following:
H5: Participants in the trustworthy condition will invest more tokens than in the untrustworthy condition.

4 METHOD

4.1 Apparatus

The application for the experiment was implemented as an immersive VR experience using Unity Engine[2] 2020.3.14f1 and ran on a VR-capable PC (Intel Core i7-10700K, Nvidia RTX 3060 8GB, 32GB RAM). The VR hardware consisted of an Oculus Quest 2 HMD and two controllers connected to the PC through the Oculus Link service. In addition, demographics and self-report questionnaires were implemented with LimeSurvey[3] 4.5.0.

Virtual Environment. The virtual environments were created with Blender[4] 2.92, and assets were downloaded from Unity Asset Store. We also implemented realistic, physically-based locomotion and interaction, so that participants could move around and interact with the virtual environment as in real life. The virtual environments of the maze task and the investment game can be seen in Figure 1B and Figure 1A, respectively.

Virtual Human. The realism of avatars' non-verbal cues often impacts the level of their social presence, their congruency with the virtual environment, and the plausibility of them being considered social interaction partners. Thus, we ensure the realism of the virtual human in several aspects. The avatar stands in an idle pose, with its head and eyes following the participants naturally. While speaking (either a real person talking through the microphone or a pre-recorded audio being played), the mouth of the avatar will move accordingly through the Oculus Lipsync[5] plugin. Our application supports avatars compatible with humanoid unity skeletons with facial blend shapes.

[2] https://unity.com/
[3] https://www.limesurvey.org/
[4] https://www.blender.org/
[5] https://developer.oculus.com/documentation/unity/audio-ovrlipsync-unity/

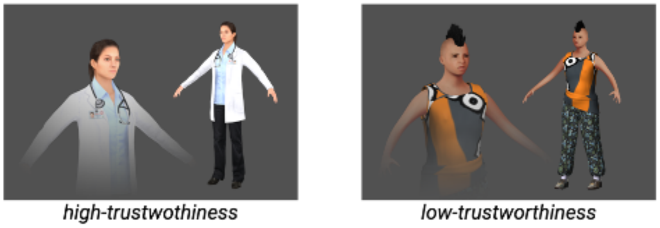


Fig. 4: The avatar chosen for the trustworthy condition (left), and the avatar chosen for the untrustworthy condition (right).

In the implementation of the validation study, the avatars are either selected from the Rocketbox avatar library [23] or created with MakeHuman[6] 1.2.0.

4.2 Pre-study: avatar selection

An online pre-study was conducted to assist in selecting the appearance of the two avatars: one for the trustworthy condition and one for the untrustworthy condition. Six avatars were selected from the Rocketbox library, and six avatars were created with MakeHuman. The trustworthiness of these avatars was evaluated and rated by 40 participants (recruited from the university's online database of study participants (SONA system)) according to the intuitive evaluation of their trustworthiness on a 7-point scale ranging from 0 - *not trustworthy at all* to 7 - *completely trustworthy*. Based on this pre-study, the most trustworthy female avatar (Score: M = 6.08, SD = 0.97) and the most untrustworthy female avatar score (Score: M = 2.30, SD = 1.02) were selected for the main experiment (Figure 4).

4.3 Participants

To reduce the impact of gender differences on the results of the experiment and given the gender distribution of the participants we were able to recruit, we decided to recruit only female participants and use only female avatars. Furthermore, our sensitivity to gender differences ensures more experimental control and reduces confounds where males and females differ in trust socialization, trust evaluation, and trust related decision making [16,56,76]. We expected a large effect size for the outcome measures based on the work of Hale et al. [26]. We recruited 70 female participants between 18 and 35 years of age without psychiatric or neurological disorders (age: M = 23.11, SD = 3.11). Participants registered to participate in the experiment via the university's SONA system. Half of the participants participated in the trustworthy condition (n = 35; age: M = 22.91, SD = 2.92), and the other half participated in the untrustworthy condition (n = 35; age: M = 23.32, SD = 3.34). Every participant was compensated with 15 euros for participating in the experiment which lasted approximately 1.5 hours.

4.4 Procedure

At the beginning of the session, participants read the study information, gave their informed consent, and filled in the pre-questionnaires (demographics and VR-related health questions). The introductory video was played after the pre-questionnaires were completed. After that, participants were guided by the experimenter to put on the HMD and enter the VR interaction tutorial.

The investment game. After the tutorial, participants started the first trial of the investment game which is an established, validated behavioural measure of trust [9]. We used a variant of the investment game in which, in the beginning, participants received 50 tokens. They can decide how many tokens out of 50 they want to invest by sending them to the trustee (see Figure 1A). The participants were told that the invested amount would be multiplied by four while being transferred. The trustee would then decide to return half of the tokens or none (participants were told they would receive feedback at the end of the experiment). The underlying assumption is that the more tokens participants invest, the higher the trust in the trustee. The investment game also serves as an additional measure for the behavioural proxies of trust in the maze task. Another trial of the investment game with the same setup was played after the maze task.

Table 2: Experimental procedure of the validation study. The two groups differ in the three VR tasks, where participants encounter different virtual humans.

Group 1 n=35 (trustworthy)	**Group 2 n=35 (untrustworthy)**
Consent, demographics, pre-questionnaires	
Introductory Video – 5 mins	
VR Interaction Tutorial	
Task 1 (VR) Investment Game	**Task 1** (VR) Investment Game
Task 2 (VR) the Maze	**Task 2** (VR) the Maze
Task 3 (VR) Investment Game	**Task 3** (VR) Investment Game
Post questionnaires	
Debriefing	

The Maze. The second VR task consisted of our version of the virtual maze task. Half of the participants (n = 35) entered the maze with the trustworthy avatar, and the other half (n = 35) entered the maze with the untrustworthy avatar. Participants will receive the same advice (if they ask for it) within each condition, and the avatar will display the same behaviour.

After completing the maze and the second trial of the investment game, participants removed the HMD to fill in the post-questionnaires. These included explicit questions about their evaluation of the avatar (humanness, friendliness, realness, helpfulness, trustworthiness) during the tasks. The Igroup Presence Questionnaire (IPQ) was used to measure presence, "the sense of being there" in the virtual environment [67]. The Social Presence Questionnaire [5] was included to examine to what extent participants perceived the avatar as a real human-being and social interaction partner. To measure simulator sickness, the Simulator Sickness Questionnaire (SSQ) was used [30]. Lastly, the General Trust Scale (GTS) [77] and KUSIV3 [6] were included to measure how participants trust others in general. The experiment concluded with a debriefing and full disclosure. Table 2 briefly summarizes the procedure of the experiment.

5 Results

5.1 Manipulation check

After each condition, participants were asked to share their experiences during the tasks on 1-7 Likert scales regarding the trustworthiness of the virtual humans. Independent sample t-tests were conducted to test the scores of these subjective measurements.

Participants in the trustworthy condition evaluated the avatar as significantly more trustworthy (M = 9.40, SD = 2.28) compared to participants in the untrustworthy condition (M = 6.46, SD = 2.86), t (68) = 4.76, p < 0.001. The results confirmed that our manipulation of trustworthiness was successful, and **H1** can be accepted.

5.2 Behavioural measures

Independent samples t-tests were conducted to test differences between behavioural measures in the trustworthy and untrustworthy conditions in both the maze and the investment game.

[6] http://www.makehumancommunity.org/

5.2.1 Advice seeking

Participants in the trustworthy condition asked for more advice (*M* = 9.91, *SD* = 2.54), compared with participants in the untrustworthy condition (*M* = 7.03, *SD* = 2.96), *t* (68) = 4.48, *p* < 0.001. A large effect size (Cohen's *d* = 1.05) was observed. Thus, **H2** can be accepted.

5.2.2 Advice following

Participants in the trustworthy condition followed more advice (*M* = 8.17, *SD* = 2.79), than those in the untrustworthy condition (*M* = 5.40, *SD* = 2.80), *t* (68) = 3.93, *p* < 0.001, with a large effect size (Cohens' *d* = 0.94).

However, the ratio of time following advice (the number of times following advice divided by the number of times asking for advice) shows no significant difference between the two conditions, *t* (68) = 1.00, *p* = 0.323. Only a tendency that participants in the trustworthy condition (*M* = 0.82, *SD* = 0.12) followed advice more often than those in the untrustworthy condition (*M* = 0.78, *SD* = 0.17) was observed. Thus, **H3** is partially rejected. On the one hand, participants in the trustworthy condition did not follow the advice more, once the advice was given; on the other hand, the number of times following advice in the trustworthy condition is indeed significantly higher, although it is highly correlated with the number of times advice was asked for.

5.2.3 Response time to execute given advice

The response times of participants in the trustworthy condition (*M* = 5.17s, *SD* = 2.14) and untrustworthy condition (*M* = 5.43s, *SD* = 2.10) were not statistically different, *t* (68) = -0.52, *p* = 0.607. Thus, participants in the trustworthy condition did not respond faster (choosing a door faster), and **H4** was rejected.

5.2.4 Investment game

The number of tokens invested between the two conditions. We conducted an independent sample t-test. In the first trial before the maze task, participants in the trustworthy condition (*M* = 28.6, *SD* = 7.51) did not invest more tokens than those in the untrustworthy condition (*M* = 27.1, *SD* = 9.84), *t* (68) = 0.71, *p* = 0.480. However, significant differences were observed, *t* (68) = 3.61, *p* < 0.001, in the second trial. Participants in the trustworthy condition (*M* = 32.5, *SD* = 8.08) invested more than those in the untrustworthy condition (*M* = 24.5, *SD* = 10.3). Thus, **H5** can be partially accepted. Only in the second trial of the investment game after the maze task, were more tokens invested in the trustworthy condition.

Differences in tokens invested between the two trials. We wanted to explore whether participants would invest more or fewer tokens after the maze task in the second trial compared to the first trial. Therefore, we conducted paired sample t-tests in both conditions.

In the trustworthy condition, significantly more tokens were invested in the second trial than in the first trial, *t* (34) = 3.30, *p* = 0.002, *mean difference* = 3.89. In the untrustworthy condition, although participants tended to invest less in the second trial, the differences are not significant, *t* (34) = -1.45, *p* = 0.157, *mean difference* = -2.63.

5.3 Exploratory measures

To verify that no bias existed in both conditions regarding participants' general propensity to trust, presence, and simulator sickness, we performed independent-sample t-tests for these measurements. There is no significant difference between the trustworthy and untrustworthy conditions in the scores of the General Trust Scale (GTS), *t* (68) = -1.67, *p* = 0.099, and KUSIV3, *t* (68) = -1.16, *p* = 0.252, indicating that participants' general propensity to trust is not biased. Similarly, no significant difference was observed for IPQ_general, *t* (68) = 0.34 , *p* = 0.736, and SSQ_total *t* (68) = -1.02, *p* = 0.313, indicating that similar levels of presence and simulator sickness were triggered in both conditions.

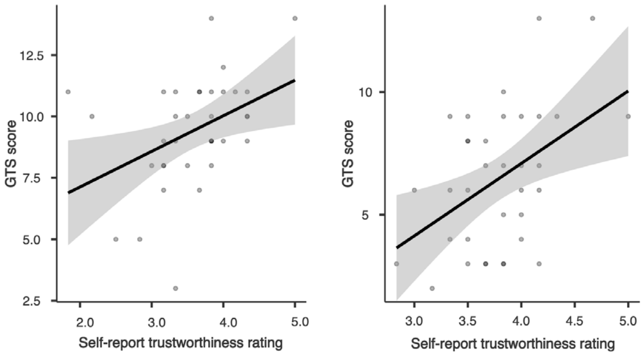


Fig. 5: Scatter plots for the correlation between the self-report trustworthiness rating and the general trust scale (GTS) score in the trustworthy condition (left) and the untrustworthy condition (right). The black lines in the plots represent the linear regression between the two variables.

We also tested whether our framing of there being another participant behind the avatar can convince our participants, and to what extent they perceive the avatar as an interaction partner. The explicit questions "Did you think you really interacted with a real person? (1-Don't believe at all; 7-Totally believe)" and "How human did you perceive the virtual human? (1-*Not human at all*; 7-*Totally human*)", as well as those from the Social Presence Scale, were included. First of all, no significant differences were observed for the three measurements between the two conditions (believed they interacted with a real person: *t* (68) = 0.81, *p* = 0.418; humanness: *t* (68) = 0.98, *p* = 0.331; social presence: *t* (68) = 0.73, *p* = 0.468), suggesting that, even though we manipulated the trustworthiness of the two avatars, including their appearances and verbal behaviour, such manipulations did not impact the social presence of the avatars, nor the effects of our framing, which exclude these factors from influencing behavioural measures. The means of the three scores (believe they interacted with a real person: *M* = 3.7, *SD* = 2.2; humanness: *M* = 4.06, *SD* = 1.46; social presence: *M* = 0.4[7], *SD* = 7.83) are all slightly higher than the medium level. Considering that they were, in fact, interacting with pre-scripted agents, such results are promising and reflect that our framing is successful. Additionally, several participants have explicitly mentioned that they really thought there was another participant after being debriefed on the truth, which also suggests the success of framing.

In addition, we tested the correlation between our subjective and behavioural measures (self-report trustworthiness ratings, advice seeking, advice following and tokens invested) with an established measure of the General Trust Scale (GTS) [77].

In the trustworthy condition, the self-report trustworthiness ratings correlate with the GTS score with a small effect, *p* = 0.040, *r* = 0.35; in the untrustworthy condition, a similar correlation was observed, *p* = 0.006, *r* = 0.45 (Figure 5). In comparison, we did not find a correlation between the GTS score and behavioural measures in the maze task. However, in the untrustworthy condition, a correlation between the tokens invested in the second trial of the investment game and the GTS score with a small effect was found, *p* = 0.035, *r* = 0.36. This may suggest that both generalized trust and specific trust influenced both the self-reports and the one-shot investment game.

Although the ratio of advice following is not sensitive to the manipulation of trustworthiness, it correlates with generalized trust (GTS), *p* = 0.024, *r* = 0.27, suggesting that those participants who generally trust others more, follow advice more often.

[7] A positive social presence score indicates that the participant believed the agent was conscious and was watching them [5].

We also noticed a positive correlation between social presence and the self-report trustworthiness ratings in both the trustworthy condition, $p = 0.005$, $r = 0.47$, and the untrustworthy condition, $p = 0.005$, $r = 0.47$. Participants with a stronger feeling of being present with a "real" person also considered the virtual humans more trustworthy, regardless of how the virtual humans looked and verbally behaved.

6 Discussion

Participants in the trustworthy condition evaluated the avatar as significantly more trustworthy than participants in the untrustworthy condition, indicating a successful experimental manipulation. Previous research suggests that visual appearance [55,70] and vocal cues [72,73] have an impact on trust; our results indicated that their manipulation does indeed influence the evaluated trustworthiness of virtual humans.

To assess interpersonal trust, we measured four behavioural proxies of trust: 1) the number of times advice was asked; 2) the number of times advice was followed; 3) the time it took participants to follow/execute the advice received; and 4) the number of tokens invested in the investment game.

In the maze task, participants asked for significantly more advice in the trustworthy condition, which could reflect the high trust evaluation of the avatar. This suggests that, in our paradigm, the number of times advice was asked for can be a good proxy for measuring trust.

In the trustworthy condition, advice given is followed slightly more often, but no statistical difference was observed. These results should not be compared to the study of Hale et al. [26] which detected significant differences in the ratio of advice following. In their design, participants can always ask for advice from two characters that give opposing suggestions, and they can only choose one to follow; in our between-subject design, participants have the choice to ask advice from only one avatar in each room. Therefore, the tendency to follow the advice might arise more from their general propensity to trust and might be moderated by their level of generalized trust, which is also in line with the positive correlation found with the GTS [77] score.

In both conditions, no statistical difference in response times (to follow advice) was observed, despite the descriptive differences. The tendency of shorter response times in the trustworthy condition could be due to feeling "safe" with the trustworthy avatar, indicating prosocial, trusting behaviour that is consistent with their subjective evaluation [17].

In the first trial of the investment game [9], no difference was detected in the two conditions. In the second trial, participants in the trustworthy condition invested significantly more tokens, which serves as an additional behavioural measure of trust. Two reasons for the results are: 1) it is possible that the manipulation of visual appearance alone is not strong enough and so does not significantly impact the perceived trustworthiness of avatars; 2) it is possible that the investment game is more sensitive to the differences in the interpersonal trust after more interaction rather than just a first impression. A significant increase in the number of tokens in the second trial compared with the first trial indicates that trust was built in the maze in the trustworthy condition. Descriptively the opposite is true for the untrustworthy condition, with no statistical difference, which could also suggest that, during the interaction with the untrustworthy avatar, the verbal behaviour and the fact that the avatar is "late" created a vulnerability in the trust relationship and lowered the expectation that the trustee would return tokens. The results also show the potential of the one-shot investment game to be a tool for assessing trust building during social interactions with virtual humans.

6.1 Advantages, limitations, and future work

The proposed variant of the virtual maze paradigm, as a behavioural measurement of interpersonal trust towards virtual humans, has the following advantages over other measures.

Firstly, self-reports are usually considered ambiguous for measuring trust [7] and tend to predict external behaviour poorly [2,44]. This may still hold, as we did not find a correlation between the self-report trustworthiness ratings and advice-seeking behaviour. Furthermore, self-reports and the one-shot investment game might be influenced by one's generalized trust level rather than only specific trust. Comparatively, the ask-seeking behaviour in the virtual maze task is not easily influenced by generalized trust.

In previous studies that measured trust towards virtual agents or avatars [8,26,52], participants were aware that they were interacting with pre-scripted agents. Their interpretation of the experiment might influence their behaviour. Therefore the trust being measured could fall into other categories of trust [49] instead of interpersonal trust. In our paradigm, participants were under the impression that they were interacting with a real person. Therefore, they were more likely to consider the virtual humans as social interaction partners, which is a prerequisite for using this paradigm in studying avatars in social VR. Furthermore, our framing method provides an alternative to multi-participant studies, which are usually costly, difficult to operate, and bring undesirable confounds.

The proposed paradigm also utilizes the advance of VR technology. It provides an immersive and interactive virtual environment that allows users to have physically-based interactions with the environment and the virtual humans (both avatars and agents). The experimental process is similar to the experience of social VR. Thus, our paradigm shows greater potential for relevant studies in VR, where fundamental VR characteristics (such as embodiment [62], virtual body ownership [69], presence [67], congruency and plausibility [38], and so on) can be further manipulated and the interplays with trust can be investigated.

It is worth noting that due to its specific setups and cover story, this paradigm is limited to experimental research and may not be suitable for measuring trust in social VR in a natural scenario (e.g., to measure whether a social VR user perceived a passing stranger avatar trustworthy or not). Instead, our variant of the virtual maze provides the academic community with a flexible foundation to build on for future investigation of trust. In the original paradigm, it is impossible to manipulate external factors that influence trust, such as cognitive load [1,66], as two virtual characters should always be included. Our modifications also provide flexibility for between-subject designs, often preferred to avoid social desirability bias [31]. Additionally, participants could potentially be more engaged in the task by randomly changing the details of each room and providing a more convincing setup (e.g., the alleged map of the maze).

Despite the advantages mentioned above, we have identified some limitations and directions for further research. In our current implementation, the social presence of the virtual humans is at a medium level. This may be because the avatars are merely standing, with no other physical movement. In future studies, more non-verbal cues (e.g., body movements and vivid facial expressions corresponding to the vocals) could be considered to increase social presence. In the current stage of social VR applications, users' avatars can already reproduce body movements and even facial expressions with additional trackers [37,46]. Meanwhile, the positive correlation between social presence and perceived trustworthiness warrants further investigation.

Another limitation is that there is no feedback regarding progress when navigating through the maze. Although this was deliberately designed to increase the uncertainty and avoid over-complicating participants' decision-making, it could potentially lead to the loss of motivation and to mindless selection. Furthermore, our paradigm has not been designed to measure the change in trust (e.g., trust building) during interactions, which could be another future direction to investigate.

Moreover, although we assume that the proposed paradigm is suitable for intelligent agents, the same observations may not still hold true, as we only considered the case of (alleged) avatars in the validation study.

7 Conclusion

Several subjective questionnaires were previously proposed to measure interpersonal trust between humans. In laboratory experiments, these tools are used and often adjusted to measure interpersonal trust between users and avatars. However, fundamental virtual reality constructs such as presence, immersion, and virtual embodiment, are neither supported nor controlled for in these tools. To date, no subjective or behavioural measurement tool (apart from Hale et al.'s novel maze paradigm [26]) was specifically designed to measure interpersonal trust towards virtual humans in virtual reality. We proposed an improved version of the maze paradigm and tested it with a validation study. Compared to Hale et al.'s paradigm, our virtual humans have more realistic social cues, and our paradigm can be used for the investigation of interpersonal trust towards either agents or avatars. In our between-subject design, participants interacted with one virtual human, thereby avoiding social desirability bias and allowing participants' behaviour to be based on their subjective evaluation of the virtual human. Our paradigm also emphasizes the immersion and interactivity of VR and considers fundamental VR characteristics. With improved details and a self-contained cover story, our design gave participants motivational incentives amid their uncertainty in the maze task. The validation study indicates that the paradigm is sensitive to the manipulation of trustworthiness. Our paradigm therefore fills the gap in the literature and suggests an in-the-moment behavioural measure of interpersonal trust for the VR community.

Figure Credits

Figure 4A was recreated from an example image of the Rocketbox avatar library [23], available at `https://github.com/microsoft/Microsoft-Rocketbox`.

Acknowledgments

This work is part of the Privacy Matters (PriMa) project. The PriMa project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860315.

References

[1] M. I. Ahmad, J. Bernotat, K. Lohan, and F. Eyssel. Trust and Cognitive Load During Human-Robot Interaction, Sept. 2019. arXiv:1909.05160 [cs]. doi: 10.48550/arXiv.1909.05160 8

[2] C. J. Armitage and J. Christian. From attitudes to behaviour: Basic and applied research on the theory of planned behaviour. *Current Psychology*, 22(3):187–195, Sept. 2003. doi: 10.1007/s12144-003-1015-5 2, 8

[3] S. Aseeri and V. Interrante. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2608–2617, May 2021. doi: 10.1109/TVCG.2021.3067783 2, 3

[4] R. Bachmann and A. Zaheer. *Handbook of Trust Research*. Edward Elgar Publishing, Jan. 2006. 2

[5] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 10(6):583–598, Dec. 2001. doi: 10.1162/105474601753272844 5, 6, 7

[6] C. Beierlein, C. J. Kemper, A. Kovaleva, and B. Rammstedt. Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen (KUSIV3). 2012. 2, 6

[7] A. Ben-Ner and F. Halldorsson. Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, 31(1):64–79, Feb. 2010. doi: 10.1016/j.joep.2009.10.001 2, 8

[8] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research*, 34(2):287–318, Apr. 2008. doi: 10.1111/j.1468-2958.2008.00322.x 2, 3, 8

[9] J. Berg, J. Dickhaut, and K. McCabe. Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1):122–142, July 1995. doi: 10.1006/game.1995.1027 2, 6, 8

[10] K. Blomqvist. The many faces of trust. *Scandinavian Journal of Management*, 13(3):271–286, Sept. 1997. doi: 10.1016/S0956-5221(97)84644-1 2

[11] D. Chan. So Why Ask Me? Are Self-Report Data Really That Bad? In *Statistical and Methodological Myths and Urban Legends*, p. 28. Routledge, 2008. 2

[12] F. Clément, M. Koenig, and P. Harris. The Ontogenesis of Trust. *Mind & Language*, 19(4):360–379, 2004. doi: 10.1111/j.0268-1064.2004.00263.x 2, 3

[13] L. L. Couch and W. H. Jones. Measuring Levels of Trust. *Journal of Research in Personality*, 31(3):319–336, Sept. 1997. doi: 10.1006/jrpe.1997.2186 2

[14] E. J. de Visser, S. S. Monfort, R. McKendrick, M. A. B. Smith, P. E. McKnight, F. Krueger, and R. Parasuraman. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22:331–349, 2016. doi: 10.1037/xap0000092 2

[15] J. D. N. Dionisio, W. G. B. Iii, and R. Gilbert. 3D Virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys*, 45(3):1–38, June 2013. doi: 10.1145/2480741.2480751 1

[16] A. H. Eagly and W. Wood. Handbook of Theories of Social Psychology: Volume 2. In *Handbook of Theories of Social Psychology: Volume 2*, pp. 458–476. SAGE Publications Ltd, London, 2012. doi: 10.4135/9781446249222 6

[17] N. Eisenberg, N. D. Eggum, and L. Di Giunta. Empathy-Related Responding: Associations with Prosocial Behavior, Aggression, and Intergroup Relations. *Social Issues and Policy Review*, 4(1):143–180, 2010. doi: 10.1111/j.1751-2409.2010.01020.x 8

[18] K. Foerster, R. Hein, S. Grafe, M. E. Latoschik, and C. Wienrich. Fostering Intercultural Competencies in Initial Teacher Education. Implementation of Educational Design Prototypes using a Social VR Environment. In *Innovate Learning Summit*, pp. 95–108. Association for the Advancement of Computing in Education (AACE), 2021. 1

[19] G. Freeman and D. Maloney. Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):239:1–239:27, 2021. doi: 10.1145/3432938 2

[20] G. Freeman, S. Zamanifard, D. Maloney, and A. Adkins. My Body, My Avatar: How People Perceive Their Avatars in Social Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–8. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3382923 2

[21] M. Freitag and R. Traunmüller. Spheres of trust: An empirical analysis of the foundations of particularised and generalised trust. *European Journal of Political Research*, 48(6):782–803, 2009. doi: 10.1111/j.1475-6765.2009.00849.x 2

[22] E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, and C. L. Soutter. Measuring Trust*. *The Quarterly Journal of Economics*, 115(3):811–846, Aug. 2000. doi: 10.1162/003355300554926 2

[23] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano, S. Orts-Escolano, V. Orvalho, L. Trutoiu, M. Wojcik, M. V. Sanchez-Vives, J. Bailenson, M. Slater, and J. Lanier. The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Frontiers in Virtual Reality*, 1, 2020. 6, 9

[24] M. A. Graber and A. D. Graber. Get Your Paws off of My Pixels: Personal Identity and Avatars as Self. *Journal of Medical Internet Research*, 12(3):e28, Sept. 2010. doi: 10.2196/jmir.1299 2

[25] A. Halbig, S. K. Babu, S. Gatter, M. E. Latoschik, K. Brukamp, and S. von Mammen. Opportunities and Challenges of Virtual Reality in Healthcare – A Domain Experts Inquiry. *Frontiers in Virtual Reality*, 3, 2022. Publisher: Frontiers. doi: 10.3389/frvir.2022.837616 1

[26] J. Hale, M. E. Payne, K. M. Taylor, D. Paoletti, and A. F. De C Hamilton. The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology*, 71(4):989–1008, Apr. 2018. doi: 10.1080/17470218.2017.1307865 2, 3, 6, 8, 9

[27] C. Johnson-George and W. C. Swap. Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, 43:1306–1317, 1982. doi: 10.1037/0022-3514.43.6.1306 2

[28] S. Jung, C. Sandor, P. J. Wisniewski, and C. E. Hughes. RealME: the influence of body and hand representations on body ownership and presence. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, pp. 3–11. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132186 3

[29] D. Kahneman, J. L. Knetsch, and R. H. Thaler. Fairness and the Assumptions of Economics. *The Journal of Business*, 59(4):S285–S300, 1986. 2

[30] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, July 1993. doi: 10.1207/s15327108ijap0303_3 6

[31] M. F. King and G. C. Bruner. Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing*, 17(2):79–103, 2000. doi: 10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0 8

[32] M. A. Koenig, F. Clément, and P. L. Harris. Trust in Testimony: Children's Use of True and False Statements. *Psychological Science*, 15(10):694–698, Oct. 2004. doi: 10.1111/j.0956-7976.2004.00742.x 3

[33] M. A. Koenig and C. H. Echols. Infants' understanding of false labeling events: the referential roles of words and the speakers who use them. *Cognition*, 87(3):179–208, Apr. 2003. doi: 10.1016/S0010-0277(03)00002-7 3

[34] M. A. Koenig and P. L. Harris. Preschoolers Mistrust Ignorant and Inaccurate Speakers. *Child Development*, 76(6):1261–1277, 2005. doi: 10.1111/j.1467-8624.2005.00849.x 3

[35] M. A. Koenig and P. L. Harris. The Basis of Epistemic Trust: Reliable Testimony or Reliable Sources? *Episteme*, 4(3):264–284, Oct. 2007. Publisher: Cambridge University Press. doi: 10.3366/E1742360007000081 3

[36] D. M. Kreps. Corporate culture and economic theory. In J. E. Alt and K. A. Shepsle, eds., *Perspectives on Positive Political Economy*, Political Economy of Institutions and Decisions, pp. 90–143. Cambridge University Press, Cambridge, 1990. doi: 10.1017/CBO9780511571657.006 2

[37] B. Lang. Researchers Show Full-body VR Tracking with Controller-mounted Cameras, May 2022. Available: https://www.roadtovr.com/full-body-vr-tracking-controller-cameras-standalone-headset-research/. [Accessed: 02-Feb-2023]. 8

[38] M. E. Latoschik and C. Wienrich. Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.694433 8

[39] J. D. Lee and K. A. See. Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1):50–80, Mar. 2004. doi: 10.1518/hfes.46.1.50_30392 2

[40] T. W. Liew, S.-M. Tan, and H. Ismail. Exploring the effects of a non-interactive talking avatar on social presence, credibility, trust, and patronage intention in an e-commerce website. *Human-centric Computing and Information Sciences*, 7(1):1–21, 2017. 1, 2

[41] J. Lin and M. E. Latoschik. Digital body, identity and privacy in social virtual reality: A systematic review. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.974652 2

[42] M. Lohle and S. Terrell. Real projects, virtual worlds: Coworkers, their avatars, and the trust conundrum. *Qualitative Report*, 19, Feb. 2014. doi: 10.46743/2160-3715/2014.1268 1

[43] M. Machneva, A. M. Evans, and O. Stavrova. Consensus and (lack of) accuracy in perceptions of avatar trustworthiness. *Computers in Human Behavior*, 126:107017, Jan. 2022. doi: 10.1016/j.chb.2021.107017 2

[44] J. McCambridge, M. d. Bruin, and J. Witton. The Effects of Demand Characteristics on Research Participant Behaviours in Non-Laboratory Settings: A Systematic Review. *PLOS ONE*, 7(6):e39116, June 2012. doi: 10.1371/journal.pone.0039116 2, 8

[45] J. McVeigh-Schultz, E. Márquez Segura, N. Merrill, and K. Isbister. What's It Mean to "Be Social" in VR?: Mapping the Social VR Design Ecology. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, pp. 289–294. ACM, Hong Kong China, May 2018. doi: 10.1145/3197391.3205451 1

[46] K. Melnick. VR Social Platform VRChat Adds Face Tracking For Avatars, 2022. Available: https://vrscout.com/news/vr-social-platform-vrchat-adds-face-tracking-for-avatars/. [Accessed: 02-Feb-2023]. 8

[47] L. C. Miller, J. H. Berg, and R. L. Archer. Openers: Individuals who elicit intimate self-disclosure. *Journal of Personality and Social Psychology*, 44:1234–1244, 1983. doi: 10.1037/0022-3514.44.6.1234 2

[48] C. Moorman, R. Deshpandé, and G. Zaltman. Factors Affecting Trust in Market Research Relationships. *Journal of Marketing*, 57(1):81–101, 1993. doi: 10.1177/002224299305700106 2

[49] R. Moradinezhad and E. T. Solovey. Investigating Trust in Interaction with Inconsistent Embodied Virtual Agents. *International Journal of Social Robotics*, Mar. 2021. doi: 10.1007/s12369-021-00747-z 2, 8

[50] R. M. Morgan and S. D. Hunt. The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, 58(3):20–38, July 1994. doi: 10.1177/002224299405800302 2

[51] C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pp. 72–78. Association for Computing Machinery, New York, NY, USA, 1994. doi: 10.1145/191666.191703 2

[52] Y. Pan and A. Steed. A Comparison of Avatar-, Video-, and Robot-Mediated Interaction on Users' Trust in Expertise. *Frontiers in Robotics and AI*, 3, Mar. 2016. doi: 10.3389/frobt.2016.00012 3, 8

[53] Y. Pan and A. Steed. The impact of self-avatars on trust and collaboration in shared virtual environments. *PLOS ONE*, 12(12):e0189078, Dec. 2017. doi: 10.1371/journal.pone.0189078 2

[54] E. Pasquini, K. Corriveau, M. Koenig, and P. Harris. Preschoolers Monitor the Relative Accuracy of Informants. *Developmental psychology*, 43:1216–26, Oct. 2007. doi: 10.1037/0012-1649.43.5.1216 3

[55] J. Peña and S.-C. Yoo. Under Pressure: Avatar Appearance and Cognitive Load Effects on Attitudes, Trustworthiness, Bidding, and Interpersonal Distance in a Virtual Store. *Presence*, 23(1):18–32, Feb. 2014. doi: 10.1162/PRES_a_00166 2, 8

[56] H. A. Rau. Trust and trustworthiness—A survey of gender differences. In *Psychology of gender differences*, Psychology research progress, pp. 205–224. Nova Science Publishers, Hauppauge, NY, US, 2012. 6

[57] T. Reeskens and M. Hooghe. Cross-cultural measurement equivalence of generalized trust. Evidence from the European Social Survey (2002 and 2004). *Social Indicators Research*, 85(3):515–532, Feb. 2008. doi: 10.1007/s11205-007-9100-z 2

[58] G. Ripka, S. Grafe, and M. E. Latoschik. Preservice teachers' encounter with social VR–exploring virtual teaching and learning processes in initial teacher education. In *SITE Interactive Conference*, pp. 549–562. Association for the Advancement of Computing in Education (AACE), 2020. 1

[59] B. G. Robbins. Measuring Generalized Trust: Two New Approaches. *Sociological Methods & Research*, 51(1):305–356, Feb. 2022. doi: 10.1177/0049124119852371 2

[60] L. A. Rosenberger, M. Naef, C. Eisenegger, and C. Lamm. Interpersonal distance adjustments after interactions with a generous and selfish trustee during a repeated trust game. *Journal of Experimental Social Psychology*, 90:104001, Sept. 2020. doi: 10.1016/j.jesp.2020.104001 2

[61] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. Latoschik. *Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities*. Mar. 2018. doi: 10.1109/VR.2018.8447550 1

[62] D. Roth and M. E. Latoschik. Construction of the Virtual Embodiment Questionnaire (VEQ). *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3556, Dec. 2020. doi: 10.1109/TVCG.2020.3023603 1, 3, 8

[63] D. Roth, K. Waldow, M. E. Latoschik, A. Fuhrmann, and G. Bente. Socially immersive avatar-based communication. In *2017 IEEE Virtual Reality (VR)*, pp. 259–260, Mar. 2017. doi: 10.1109/VR.2017.7892275 1

[64] J. B. Rotter. A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35(4):651–665, 1967. doi: 10.1111/j.1467-6494.1967.tb01454.x 2

[65] D. Salanitri, G. Lawson, and B. Waterfield. The Relationship Between Presence and Trust in Virtual Reality. In *Proceedings of the European Conference on Cognitive Ergonomics*, ECCE '16, pp. 1–4. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2970930.2970947 3

[66] K. Samson and P. Kostyszyn. Effects of Cognitive Load on Trusting Behavior – An Experiment Using the Trust Game. *PLOS ONE*, 10(5):e0127680, May 2015. doi: 10.1371/journal.pone.0127680 8

[67] T. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, 15:69–71, Apr. 2003. doi: 10.1026//1617-6383.15.2.69 6, 8

[68] D. Shao and I.-J. Lee. Acceptance and Influencing Factors of Social Virtual Reality in the Urban Elderly. *Sustainability*, 12(22):9345, Jan. 2020. doi: 10.3390/su12229345 1

[69] M. Slater. Towards a digital body: The virtual arm illusion. *Frontiers in Human Neuroscience*, 2, 2008. doi: 10.3389/neuro.09.006.2008 8

[70] A. Surprenant. Measuring Trust In Virtual Worlds: Avatar-mediated Self-disclosure. *Electronic Theses and Dissertations*, Jan. 2012. 2, 8

[71] D. Thalmann. The Role of Virtual Humans in Virtual Environment Technology and Interfaces. In R. A. Earnshaw, R. A. Guedj, A. v. Dam, and J. A. Vince, eds., *Frontiers of Human-Centered Computing, Online Communities and Virtual Environments*, pp. 27–38. Springer, London, 2001. doi: 10.1007/978-1-4471-0259-5_3 1

[72] I. Torre. *The impact of voice on trust attributions*. Thesis, University of Plymouth, 2017. Accepted: 2017-08-24T14:30:41Z. 8

[73] E. Tsankova, A. Aubrey, E. Krumhuber, G. Möllering, A. Kappas, D. Marshall, and P. Rosin. *Facial and Vocal Cues in Perceptions of Trustworthiness*, vol. 7729. Jan. 2013. doi: 10.1007/978-3-642-37484-5_26 8

[74] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Transcations on Visualization and Computer Graphics*, 24(4):10, 2018. doi: 10.1109/TVCG.2018.2794629 3

[75] M. Wang. Social VR : A New Form of Social Communication in the Future or a Beautiful Illusion? *Journal of Physics: Conference Series*, 1518:012032, Apr. 2020. doi: 10.1088/1742-6596/1518/1/012032 1

[76] Y. Wu, A. S. M. Hall, S. Siehl, J. Grafman, and F. Krueger. Neural Signatures of Gender Differences in Interpersonal Trust. *Frontiers in Human Neuroscience*, 14, 2020. 6

[77] T. Yamagishi and M. Yamagishi. Trust and commitment in the United States and Japan. *Motivation and Emotion*, 18(2):129–166, June 1994. doi: 10.1007/BF02249397 2, 6, 7, 8

[78] C. A. Zanbaka, A. C. Ulinski, P. Goolkasian, and L. F. Hodges. Social responses to virtual humans: implications for future interface design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pp. 1561–1570. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1240624.1240861 2