Trust towards Virtual Humans in Immersive Virtual Reality and Influencing Factors

A Doctoral thesis submitted to the

Faculty of Human Sciences

at

Julius-Maximilians-Universität Würzburg

submitted by

Johrine Cronjé

from

Bloemfontein, South Africa

2023

Submitted on: ________________________________

Members of the Promotionskomitee:

Prof. Paul Pauli, Institut für Psychologie, Universität Würzburg

Prof. Marc Erich Latoschik, Informatik IX, Universität Würzburg

Prof. Dr. Carolin Wienrich, Universität Würzburg

Date of Public Defence: ____________________________

Date of Certificates: _______________________________

“Trust opens up new and unimagined possibilities.”

~ Robert Solomon

"Trust is the bridge that allows us to embrace the potential of virtual humans in virtual reality, transforming a simulated experience into a genuine connection."

~ Chat GPT

“The use of technology is a human task.”

~ Daniel Shore

Acknowledgements

First and foremost, I would like to thank Prof. Dr. Paul Pauli and Prof. Dr. Marc Erich Latoschik for allowing me to pursue this exciting research into the evaluation and acceptance of social interaction partners in virtual reality. Sincere appreciation to Dr. Ivo Käthner; I am immensely grateful for your guidance, patience, expertise, time, and unwavering support throughout this journey. Your valuable insights, constructive feedback, and dedication have shaped this research and pushed me to strive for excellence.

I am grateful to my fellow PriMa project Ph.D. students, colleagues of the Psychology Chair, Human-Computer- Interaction Chair, and friends from the Ph.D. Statistics group who have provided intellectual discussions, shared their expertise and offered assistance and distractions when needed. Your camaraderie and collaboration have made this Ph.D. journey both enriching and enjoyable. Thanks to my collegue and friend Jinghuai Lin for your companionship, critical thinking, and enriching perspective on social VR. A special thank you to Dr. Octávia Madeira for being my friend, listening, sharing, and investing time in proofreading drafts; I am so grateful and indebted to you.

Thank you to the Marie Curie-Skłodowska EU Horizon: PriMa project, Prof. Dr. Raymond Veldhuis, and Dr. Bridgette Connell, PriMa secondment partners, who have provided resources to conduct my research, opportunities to gain experience and unforgettable memories. My deepest gratitude goes to my family and friends for their unconditional love, understanding, and support throughout this long and challenging journey. Your encouragement and belief in my abilities have been a constant source of strength. To my loving husband, Dr. Andrew Soane, you have been my rock, inspiration, and balance from the beginning. Thank you for listening, comforting, encouraging, and being with me emotionally, physically, and intellectually through this journey. You still are my greatest gift and best friend!

Completing this Ph.D. thesis would not have been possible without these individuals and institutions' collective support and encouragement. Thank you for participating in this significant milestone in my academic and personal growth.

Abstract

Virtual humans (VHs) hold immense potential for collaboration in social virtual reality (VR). As VR technology advances, it's vital to assess the psychological effects on VH trust and user privacy to build meaningful social interactions in VR. In social VR, users must be able to trust the VHs they interact with as they navigate through socio-cultural activities. The evaluation of trustworthiness in VHs profoundly impacts interaction quality and user willingness to engage. Conversely, untrustworthy VHs can harm user experiences, privacy, and VR engagement. To address this, we conducted immersive VR studies, exploring how psychological factors influence user's VH trust evaluation under various psychological conditions. This research is pivotal for developing strategies to enhance user privacy, establish secure VR environments, and create a foundation of trust that supports immersive socio-cultural experiences in VR.

To date, there are no established interpersonal trust measurement tools specifically for VHs in VR. In study 1 (the *familiarity study*) of the current thesis the VR-adjusted version of the social conditioned place preference paradigm (SCPP) by Kiser et al., (2022) was identified as a potential trust measurement tool. We tested whether the familiarity of a VH influenced trust as measured with the SCPP paradigm and other self-defined outcome measures, in a Computer Augmented Virtual Environment (CAVE). The CAVE is a VR system that combines immersive VR with real-world elements. It consists of a room-sized space where the walls are used as projection screens to display virtual scenes and objects. In this within - subject design (n = 20), half of the participants were familiarized with one VH and tasked to explore and interact in a realistic looking virtual art museum environment. The participant's evaluation of the VH's trustworthiness was measured as well as their subsequent trust behaviours. Results revealed no significant differences in the evaluation of the VH's trustworthiness nor any behavioural differences between conditions. The findings of the impact of a VH's familiarity on trust is inconclusive due to the major limitations of the paradigm. We concluded that the SCPP paradigm needs further validation and the proposed proxies of trust need to be re-evaluated. The findings were considered in the following study.

The virtual maze paradigm design of Hale, (2018) was identified as a potential trust measurement tool, however several limitations are associated with its use to measure trust in VR. In study 2 (a *validation study*), improvements were made to the virtual maze paradigm of Hale, (2018) and a variant of this paradigm was implemented. We conducted a validation study with 70 participants in a between-subject design with VH trustworthiness as the between-subject factor. Participants wore a head-mounted display (HMD), to deliver an immersive VR

experience. In our version of the virtual maze, it was the task of the users (the trustors) to navigate through a maze in VR, where they could interact with a VH (the trustee). They could choose to ask for advice and follow the advice from the VH if they wanted to. The number of times participants asked and followed advice and the time it took to respond to the given advice served as behavioural proxies/measures of trust. The two conditions (trustworthy vs. untrustworthy) did not differ in the content of the advice but in the appearance, tone of voice and engagement of the trustees (allegedly an avatar controlled by other participants). Results indicated that the experimental manipulation was successful, as participants rated the VH as more trustworthy in the trustworthy condition compared with the VH in the untrustworthy condition. Importantly, this manipulation affected the trust behaviour of participants, who, in the trustworthy condition, asked for advice and followed advice more often, indicating that the paradigm is sensitive to differences in VH's trustworthiness. Thus, our paradigm can be used to measure differences in interpersonal trust towards VHs and may serve as a valuable research tool for researchers who study trust in VR. Therefore, study 2 fills the gap in the literature, for an interpersonal trust measurement tool specifically for VHs in VR.

Two experimental studies, with a sample size of 50 participants each, utilized the virtual maze paradigm where participants entered 12 rooms under different conditions. We examined the influence of cognitive load (CL) on trust towards VH in VR in study 3 (*Cognitive load study*), and the influence of emotional affect (*Emotional affect study*) on trust towards VH in VR in study 4 (*EA study*). In both studies, we assessed participant's evaluation of a VH's trustworthiness, along with three behavioural indicators of trust in the maze task: 1) frequency of advice asked, 2) frequency of advice followed, and 3) the time taken by participants to execute the received advice. In study 3, the CL was manipulated with the auditory 1-back task in the high cognitive load condition (HCL). In study 4, the Autobiographical Emotional Memory Task (AEMT) was used to manipulate the EA of participants in the negative emotional affect (NEA) condition. As an additional manipulation, while participants were immersed in VR, they were exposed to 12 negative pictures and sounds that was presented simultaneously to strengthen the initial manipulation. The manipulation of the within-subject factors (CL and EA) was successful in both studies, as significant differences between conditions were observed in both studies (higher CL in the HCL condition and a more negative EA in the NEA condition). However, only CL influenced participant's evaluation of the VH's trustworthiness. The VH were evaluated as significantly more trustworthy after the HCL condition. Despite the difference in trust evaluation, there was no difference in advice asking or following. Participants in study 4 asked and followed advice due to their trust in the VH and asked and followed advice equally often in both conditions. Importantly, significant differences were observed in the participants response times in both studies. In study 3 during the HCL condition participants followed advice quicker. The order in which the conditions were presented influenced the experience of CL. Participants experienced higher levels of CL and responded to advice significantly faster when low cognitive load (LCL) was presented as the first condition compared with LCL as the second condition. In study 4 participants in the NEA condition followed advice slower similar to the findings of study 3. The order in which the conditions were presented had a significant effect on the EA. Participants asked and followed advice less when the NEA condition was presented first compared with when it is presented second. Possible explanations for the findings are discussed in the thesis.

Overall, this thesis offers a novel tool for trust measurement (the virtual maze paradigm) and contributes to understanding the role of psychological factors in trust towards virtual humans in virtual reality.

Contents

List of Tables

List of Figures

Abbreviations

AEMT	**A**utobiographical **E**motional **M**emory **T**ask
BFI	**B**ig **F**ive **I**nventory
CASA	**C**omputers **A**re **S**ocial **A**ctors
CAVE	**C**omputer **A**utomatic **V**irtual **E**nvironment
CL	**C**ognitive **L**oad
COVID-19	**C**oronavirus **D**isease
CPP	**C**ondition **P**lace **P**reference
EA	**E**motional **A**ffect
EU	**E**uropean **U**nion
ESS	**E**uropean **S**ocial **S**urvey
GTS	**G**eneral **T**rust **S**cale
HCL	**H**igh **C**ognitive **L**oad
HMD	**H**ead- **M**ounted **D**isplay
Hz	**H**ert**z**
IADS-E	**A**ffective **A**uditory **S**timulus **D**atabase
IAPS	**A**ffective **P**icture **S**ystem
IPQ	**I**group **P**resence **Q**uestionnaire
IT	**I**nterpersonal **T**rust
ITS	**I**nterpersonal **T**rust **S**cale
IVR	**I**mmersive **V**irtual **R**eality
KUSIV3	Das **K**urze zwischenmenschliche **V**ertrauen
LCL	**L**ow **C**ognitive **L**oad
LED	**L**ight-**E**mitting **D**iode
NEA	**N**egative **E**motional **A**ffect
PANAS	**P**ositive and **N**egative **A**ffect **S**chedule
PEA	**P**ositive **E**motional **A**ffect
PHN	**P**hilosophies of **H**uman **N**ature **S**cale
PRIMA	**PRI**vacy **MA**tters
PTS	The **P**ropensity to **T**rust **S**urvey
SAM	**S**elf-**A**ssessment **M**anikin Questionnaire
SCPP	**S**ocial **C**onditioned **P**lace **P**reference
SDI	**S**elf-**D**isclosure **I**ndex

SFT	**S**tranger-**F**ace **T**rust
SIP	**S**ocial **I**nteraction **P**artner
SITS	**S**pecific **I**nterpersonal **T**rust **S**cale
SCPP	**S**ocial **C**onditioned **P**lace **P**reference
SSQ	**S**imulator **S**ickness **Q**uestionnaire
STAI	**S**tate-**T**rait **A**nxiety **I**nventory
SVR	**S**ocial **V**irtual **R**eality
T	**T**rustworthy
TG	**T**rust **G**ame
U	**U**ntrustworthy
VH	**V**irtual **H**uman
VR	**V**irtual **R**eality
XR	**E**xtended **R**ealities

1. Introduction

This PhD thesis is part of a larger research project called Privacy Matters (PriMa). The project aims to analyze and mitigate privacy risks in the digitalizing European society. For example, the loss of privacy due to the increase in recognition technologies that make inferences from biometric data about an individual's emotional state, age, gender, etc., or possible privacy and security threats to virtual humans (VHs) and virtual reality (VR) environments such as social engineering or identity infringement. The primary objective of the project is to "gain a comprehensive understanding of the multidisciplinary nature of privacy protection in a digitalised society and to provide solutions that address this important societal challenge"(PriMa – Privacy Matters, 2018).

This thesis focuses on trust towards VHs in immersive VR, where users can interact and engage with realistic virtual environments. Social VR, is an application of immersive VR, where users can interact and collaborate with other users in the form of VHs (Lin & Latoschik, 2022). Trust is a fundamental component that inspires social interactions and fosters a sense of safety and predictability in environments (e.g., Rotter, 1967). Trust in the virtual environment and towards virtual social interaction partners (SIP) is a critical factor in ensuring user comfort, engagement, and acceptance of the technology, fostering meaningful connections and effective collaboration (e.g., Mystakidis et al., 2021; Salanitri et al., 2016; Scavarelli et al., 2021). Therefore, this thesis explores the psychological factors that influence trust toward VHs in VR. Trust in social VR interactions mirrors the importance of trust in ordinary day-to-day interactions. As it is in face-to-face encounters, users evaluate the trustworthiness of others and make decisions based on this assessment, similar to ordinary day-to-day interactions (Mayer et al., 1995). However, little research has been conducted to understand how psychological factors influence trust towards VHs in VR. Various experimental studies focus on the impact of VH's appearance on trust, for example, VH's age (Lee et al., 2018), the projection of users in virtual worlds (Lohle & Terrell, 2014), self-VHs or no self-VHs (Pan & Steed, 2017), VH's facial appearance (Machneva et al., 2022), VH voice (Siehl et al., 2022), robot vs human-like VH (George et al., 2018) to name a few. The psychological factors that influence user's evaluation of the VHs trust in social VR has not gained much attention in the academic community, rather VR was used as a research tool to investigate the impact of colour and emotion on trust (Felnhofer et al., 2015; Kim & Lee, 2022), cognitive load (Johnson-Glenberg, 2018) or using VR to raise awareness of social engineering (Jansen & Fischbach, 2020). But the link to how these psychological factors influence trust toward VHs is still missing.

Therefore, this research contributes towards the understanding of psychological factors that influence user's evaluation of VH's trust in social VR and can be used to enhance the user experience in virtual environments.

1.1. Outline of Thesis

The aim of this dissertation project was twofold. The first research goal was to identify and use a behavioural trust measurement tool, to measure interpersonal trust towards VHs in VR. The second goal was to use this tool to study psychological factors that influence trust towards VHs in VR. The present thesis is structured as follow: Chapter 2 will introduce the theoretical background consisting of the construct and conceptualization of trust, factors influencing trust, and the measurement of trust (2.1), an introduction to VR, the fundamental VR characteristics, trust in virtual social interactions, and VR as a research tool (2.2) and finally the research objectives (2.3). Interpersonal trust toward VHs was investigated using four VR studies. The empirical studies are described in Chapters 3,4,5, and 6. In study 1, described in Chapter 3, we tested whether the familiarity of a VH influenced trust. We constructed a virtual maze paradigm in study 2, described in Chapter 4, that took fundamental VR characteristics into account and validated its sensitivity to the manipulation of VH trustworthiness. We concluded that this virtual maze paradigm is suitable for measuring trust towards VHs in VR. The virtual maze was used in the study 3 (cognitive load study: Chapter 5) and study 4 (emotional affect study: Chapter 6) to investigate the influence of psychological factors on trust evaluation and consequent trust behaviours toward a VH in VR. The final chapter of the thesis (Chapter 7) provides a summary, discussion, highlights the limitations of the studies, and outlines the overall contribution of the thesis to the field. Additionally, suggestions for future research directions are provided.

2. Theoretical background

2.1. Trust: Interpersonal Trust

2.1.1. Construct and Categorization

Trust has been extensively studied in different contexts, such as: social sciences, behavioural economics, human-computer-interaction, education, and neuroscience. (Alós-Ferrer & Farolfi, 2019; Babel et al., 2021; Bee et al., 2011; Ben-Ner & Halldorsson, 2010; Caldwell & Clapham, 2003; Chiou et al., 2020; Iacono & Weisband, 1997; Moorman et al., 1993; Morgan & Hunt, 1994). There are various forms of trust, for example: trust can be person-to-person, person-to-business, person-to-robot, person-to-computer, and person-to-automation (Moradinezhad & Solovey, 2021). Due to the multi-disciplinary interest in trust research, there are many categories of trust: institutional trust, social trust, political-based trust, cognitive-based trust, emotional-based trust, affective-based trust, swift trust, intrapersonal trust, and interpersonal trust. Given the increasing prevalence of social interaction within virtual reality (VR), it becomes crucial to inquire about the specific category that encompasses the trust people develop toward virtual humans in this context (Lin et al., 2023). The research described in the current thesis focuses on interpersonal trust towards a specific SIP in social VR.

The type of trust we will focus on in this research refers to the dyadic relationship between one person and another *specific* SIP. We adopt the definition of *interpersonal trust* as "the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p. 51). Interpersonal trust is the foundation for all social situations that demand cooperation and interdependence (e.g., loaning money, visiting a medical doctor, or sharing personal information) (Johnson-George & Swap, 1982). In social interactions, the risk associated with vulnerability or dependence must be weighed against the potential for positive outcomes (Johnson-George & Swap, 1982). The fundamental nature of this trust dilemma is the inherent element of risk in social interactions that prevents any guarantee of a satisfactory outcome, despite critical assessments of the intentions, competence, and motives of the individuals involved. Trust is a crucial factor in the adaption and efficiency of new processes. Without trust, people may be hesitant to fully embrace new technologies or processes, limiting their potential benefits and hindering progress. New processes and change are usually accompanied by high levels of uncertainty, vulnerability, and risk (Rotter, 1967). Empirical studies have shown that social interactions are based on interpersonal trust (Bente et al., 2014; Caldwell & Clapham, 2003) and that no lasting relationship can be established or maintained without it (Moorman et al., 1993; Morgan & Hunt, 1994).

General trust is the stable concept of trust over time, context, and across relevant entities which influence trust toward a specific entity (Couch & Jones, 1997; Rotter, 1967; Siegrist et al., 2005; Yamagishi, 2011; Yamagishi & Yamagishi, 1994). It inclines optimistic assumptions, an expectation of goodwill, and benign intent about any entity to trust, independent of familiarity with it. It is often seen as a trait and represents the disposition to trust in any context, e.g., interaction partners, information systems, experts and expertise, institutions, and governments (Couch & Jones, 1997; Rotter, 1967; Siegrist et al., 2005; Yamagishi, 2011; Yamagishi & Yamagishi, 1994). General trust is, therefore, essential for interactions and trust in entities when information about them is limited (Yamagishi & Yamagishi, 1994).

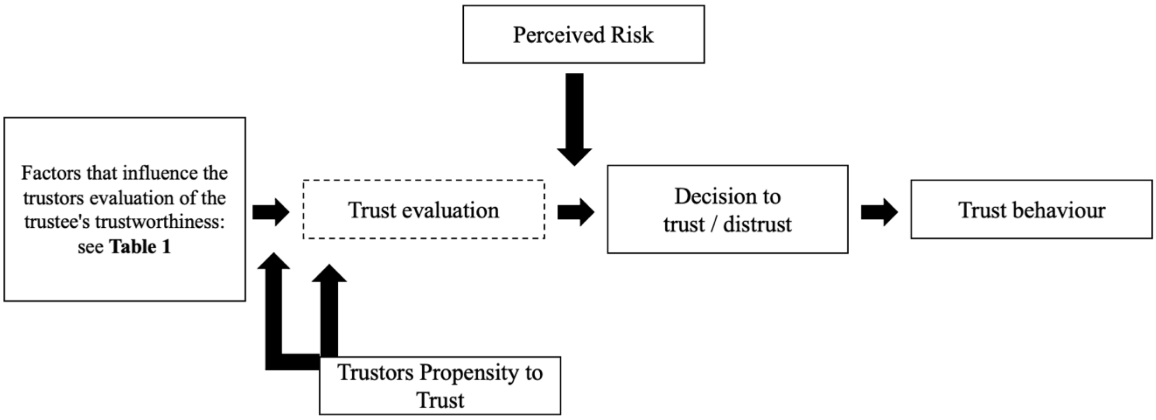
Besides the general tendency to trust an interaction partner, *specific trust* towards a specific person is important. While the willingness to take a risk to trust another is common to all trust situations, the willingness to behave in a trustful way in any given situation will be determined by various factors. For example, the individual trusted to repair a car may not be trusted to feed the cat while the owners are on vacation, and the trusted pet sitter may not be the chosen confidant for intimate self-disclosures. Specific trust is thus the trust towards a specific other in a specific relationship and context: a close relationship or a specific stranger (Johnson-George & Swap, 1982).

To assess trust on a diverse basis, the approach of the current thesis included different aspects of trust, including behavioural and perceptional aspects, as well as situational assessment and measurement of trust (e.g., Balliet & Van Lange, 2013; Hancock et al., 2011; Johnson & Mislin, 2011; Müller et al., 2020; Thielmann & Hilbig, 2014, 2015; Yamagishi, 2011).

2.1.2. Models of Trust

Theoretical models on interpersonal trust (Mayer et al., 1995) and trust in automation (Lee & See, 2004) demonstrate that regardless of whether a human or computer-controlled agent offers assistance, the fundamental elements of trust processes are *factors* that influence the evaluation of trustworthiness, *trust evaluation*, the *decision to trust* or *not* and trust-related *behaviour* (Langer et al., 2023) (see Figure 1).

Figure 1: Trust process (figure based on Mayer, Davis and Schoorman 1995)



Note: This figure shows the process of trust. Initially, a conscious or unconscious evaluation of a social interaction partner's trustworthiness is made. The evaluation is influenced by psychological and environmental factors and the trustor's propensity to trust people in general. The decision to trust or distrust a social interaction partner is influenced by the perceived risk associated with the decision. Ultimately, trust behaviour follows as a consequence of the trust decision. The trust process is dynamic and repeats over time.

The trustee's trustworthiness is evaluated based on conscious or unconscious psychological factors (Table 1 lists a few). The trustor's evaluation of the trustee's trustworthiness is further influenced by the degree to which the trustor expects the promises or statements of the trustor to be dependable and trustworthy (Mayer et al., 1995). The decision to trust or distrust the trustor is influenced by the perceived risk associated with the decision (Alarcon et al., 2018). The consequent behaviour of trust is the behavioural outcome of trust, for example, taking a risk in a relationship or following the given advice (Lee & See, 2004; Mayer et al., 1995). This suggests that trustors are willing to take on the risk of potential task failure or receiving unsound advice (Mayer et al., 1995). Depending on the resulting outcome, such as the quality of work produced or the accuracy of the given advice, trustors may reassess the trustworthiness of the trustee, thus initiating a new cycle in the trust process (Lee & See, 2004; Mayer et al., 1995).

Trust is a dynamic process that changes over time (Glikson & Woolley, 2020). Therefore, certain levels of trustworthiness, trust, and trust-related behaviour may increase over time if trustors evaluate the trustee's trustworthiness as sufficient (Lee & See, 2004; Mayer et al., 1995). Nevertheless, relying on someone or something always carries the risk of unfulfilled expectations and trust violations (e.g., poor quality work outcomes), which can decrease trustworthiness, trust, and trust-related behaviour (Kim et al., 2006). However, trust can be repaired (e.g., through excuses) and rebuilt, leaving a positive impact on evaluations of trustworthiness, trust, and trust-related behaviour (De Visser et al., 2018; Tomlinson & Mryer, 2009). See Table 1 for a summary of the factors of interest that influence trust evaluation.

Table 1: Factors of Interest that Influence Trust Evaluation.

Factors influencing trust (not limited to)	**References of experimental studies** (not limited to)
Ability, Integrity, Benevolence	Mayer et al. (1995)
Personality	Mattarozzi et al. (2015); Thielmann & Hilbig (2014)
Gender	Lee (2008); Mattarozzi et al. (2015); Wu et al. (2020)
Emotional expressions	Oosterhof & Todorov (2008); Pfaller et al. (2021)
Perceived attractiveness	Principe & Langlois (2013); Sofer et al. (2015)
Face familiarity	Zebrowitz et al. (2007)
Body language and clothing	Greenlees et al. (2005)
Friendliness	Brown et al. (2004)
Self-presentation in a specific context	Johnson et al. (2014)
Cognitive load	Duffy & Smith (2012); Peña & Yoo (2014); Samson & Kostyszyn (2015)
Emotional state	Bagneux et al. (2012); Li et al. (2019); Myers & Tingley (2016)
Affective processing	Jones (2019)
Impulsivity	Burnett Heyes et al. (2012); Eben et al. (2020)
Environment	Jones (2019)
Situation	Marsh & Dibben (2005)
Propensity to trust	Alarcon et al. (2018); Robbins (2022)
The specific person	Johnson-George & Swap (1982); Rotter (1967)
Physical appearance	Olivola & Todorov (2010); Peña & Yoo (2014)
Verbal and non-verbal behaviour	Liew et al. (2017); Olivola & Todorov (2010)
Reliability	Koenig & Harris (2007)
Consistent behaviour	Moradinezhad & Solovey (2021)
Competence	Kim et al. (2006)
Presence	Bente et al. (2004); Salanitri et al. (2016)
Humanness	de Visser (2012); Tripp et al. (2011)

Note: The factors that influence interpersonal trust are not limited to the factors listed in this table. The listed factors and supported references of experimental studies that investigate how the factors influence trust were selected based on their relevance to this project.

2.1.3. Measurement of Interpersonal Trust

Parts of the following section have already been published as:

J. Cronjé, J. Lin, I. Käthner, P. Pauli and M. E. Latoschik, "Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2401-2411, May 2023, doi: 10.1109/TVCG.2023.3247095.

Measurements of trust in experimental research are generally among two categories: subjective and objective measures, as summarized in Lin et al (2023).

Subjective measurements of trust are primarily self-report questionnaires, the predominant method to measure trust across different domains in psychology, neuroscience, sociology, and organizational science (Hale et al., 2018). These include the "Interpersonal Trust scale" (ITS) developed by Rotter (1967), the shortened version of the ITS, "KUSIV3" by Beierlein et al., (2012), and other alternatives such as the "General Trust Scale" (GTS) by Yamagishi & Yamagishi (1994). In addition, other relevant scales, including the "Self-disclosure Index" (SDI: Miller et al., (1983) and the "European Social Survey" (ESS: Reeskens & Hooghe (2008), are often used as additional measures of trust. However, most of these questionnaires only measure generalized trust (Couch & Jones, 1997) - a reflection of how much a person trusts others in general (Hale et al., 2018), rather than specific trust—trust towards a specific person either to people with close relationships or strangers (Hale et al., 2018). Robbins (2022), on the other hand, constructed the "Stranger-Face Trust" (SFT) questionnaire that aims to measure trust in specific strangers and particular matters.

As Chan (2008) has pointed out, most self-report methods can reflect internal feelings less accurately compared with objective measures. Subjective trust measurements are considered poor predictors of external behaviour (Armitage & Christian, 2003; Glaeser et al., 2000; McCambridge et al., 2012) and may not be ideal for measuring specific trust as there can be multiple interpretations of items and trust (Ben-Ner & Halldorsson, 2010). Thus, objective measurements are sometimes preferred. Objective methods often use behavioural clues during social interactions as proxies of trust. The *Trust Game* (Kreps, 1990) is one of the most popular and established trust measures 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 increased 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 for 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 that no reward will be returned. Similar ideas are adopted by a variant of *the Trust Game* or similar paradigms, including the *Dictator Game* (Kahneman et al., 1986), and the *Investment Game* (Berg et al., 1995), to name a few. Additionally, research indicates that interpersonal distance between SIPs, advice-seeking behaviour, and the duration of mutual gaze (Aseeri & Interrante, 2021; Clément et al., 2004; Hale et al., 2018; Peña & Yoo, 2014; Rosenberger et al., 2020) can be indicators of trust. A more comprehensive review of the measurements of interpersonal trust can be found in Appendix A, Table 13, and Table 14.

2.2. Virtual Reality

The term *virtual reality* was coined in 1987 by Jaron Lanier, whose research and engineering contributed to several products of the nascent VR industry (Gigante, 1993). VR was defined as "the illusion of participation in a synthetic environment rather than external observation of such an environment. VR relies on three-dimensional (3D), stereoscopic, head-tracked displays, hand/body tracking, and binaural sound. In VR an immersive, multisensory experience is possible." (Gigante, 1993, p.1). VR can be non-immersive or immersive: *in non-immersive VR*, virtual environments are displayed on a computer monitor or smartphone, where users can interact with the virtual environment using input devices like a keyboard, mouse, or touchscreen. In *Immersive VR,* users experience being physically present in a non-physical world. This experience is created through a VR system that displays the sensory data of a user's immediate surroundings from their viewpoint within the VR environment. Part of the user's immediate surroundings includes the representation of the user's body, from the unique position and orientation place of the participant's viewpoint within the environment (Slater & Usoh, 1999). Users usually use VR controllers to interact with the virtual environment. While immersive VR refers to the broader concept of VR experiences that provide a high level of immersion and presence, *social VR* is a subsection of immersive VR that specifically focuses on creating virtual environments where users can interact, communicate, and engage with each other in real time, simulating social interactions within the virtual space. Furthermore, social VR can be described as a simulation of face-to-face communication, synchronous dialogue (via audio intercommunication), and synchronous body movements embodied in an anthropomorphic 3D model (virtual human) (Blackwell et al., 2019).

When referring to VR in this thesis, we are referring to immersive social virtual reality, however the umbrella term VR will be used. The user's experience in VR is influenced by fundamental constructs of VR, namely *presence, plausibility,* and *congruence,* which are essential aspects of extended realities (XR) (Latoschik & Wienrich, 2022). *Presence* is an important emergent property of an immersive system and refers to the participant's sense of "being there" in the world created by the VR system (Latoschik et al., 2017; Slater & Usoh, 1999). An example of presence in VR is when a user is wearing a VR headset and feels completely immersed in a simulated environment, to the extent that they forget about their physical surroundings and genuinely feel as if they are present and interacting with the virtual world. This sense of "being there" and the feeling of presence can enhance the overall experience and engagement in VR. *Plausibility* refers to a subjective state or condition that arises during an experience based on the evaluation and alignment of information processed by the sensory, perceptual, and cognitive layers (Latoschik & Wienrich, 2022). For example, in VR, plausibility can be illustrated by a scenario where a user explores a virtual forest. The plausibility of the experience would be enhanced if the visual, auditory, and tactile cues align with the user's expectations and create a sense of immersion and believability, leading to a subjective state of feeling present in the virtual environment. *Congruence* denotes the objective agreement between the information that has been processed and the information that was expected across the sensory, perceptual, and cognitive layers (Latoschik & Wienrich, 2022). An example of congruence in VR could be when a user interacts with a virtual object that behaves and responds in a manner consistent with their expectations and real-world physics. For instance, if a user tries to push a virtual ball and it rolls and reacts as expected, matching their sensory and cognitive understanding of how objects should behave, it will demonstrate congruence between their anticipated and perceived information in the virtual environment. These key fundamental VR constructs of presence, plausibility, and congruence play significant roles in shaping the user's experience in VR. These constructs work together to enhance the user's engagement, believability, and overall immersion in the virtual environment. By creating a strong sense of presence, ensuring plausibility, and maintaining congruence, VR can provide users with compelling and realistic experiences that bridge the gap between the physical and virtual realms. For example, social VR shows great potential for collaboration with other users in VR in the form of VHs (Lin & Latoschik, 2022).

The Perceived Differences between Humans and Virtual Humans

People, in general, distinguish between the "real world" and "virtual reality." The virtual world is a representation, while the real world is not (Wang, 2020). Innovations in VR promote more effective dyadic (i.e. 2-person) and n-person interactions (Blascovich, 2002; Normand et al., 1999). In the development of VR applications, *virtual humans* play an important role. In immersive social VR, users can meet and interact virtually in the form of a VH. A VH is a computer-generated three-dimensional digital representation of the user, which can be human-like or cartoon-like in appearance and acts like a real human (Bombari et al., 2015; Latoschik et al., 2017). Virtual SIPs can either be *avatars* (VH's controlled by humans) or *agents* (VHs controlled by algorithms) (Blascovich et al., 2002; Latoschik et al., 2017; Pan & Hamilton, 2018). For the purpose of this thesis, when speaking about SIPs in social VR, the generic term: virtual human will be used. To clarify, a VH in social VR in this thesis refers to a digital representation of a human-like character or avatar that users can encounter and interact with within the virtual environment. These VHs are designed to simulate human appearance, behaviour, and communication to create a sense of social presence and engagement. They can exhibit realistic movements, gestures, and facial expressions, and may be programmed with artificial intelligence to engage in conversations and respond to user interactions.

As social VR grows in popularity, the use of VHs that look like users in the real world is encouraged (Kyrlitsias & Michael-Grigoriou, 2022). VHs have various applications: for instance, in social VR (McVeigh-Schultz et al., 2018), VHs can be utilized as the digital bodies of users, allowing users to be immersed in cyberspace to communicate, interact, and collaborate (Roth et al., 2017). In many scenarios, such as online healthcare or financing, VHs can act as AI assistants, enhancing their social presence (Liew et al., 2017) and potentially increasing users' acceptance. The development of VHs, especially those with a realistic appearance, can serve the purpose of enhancing virtual embodiment (Roth et al., 2018), recreating the real world (Thalmann, 2001), and making the virtual world an alternative realm for human socio-cultural activities (Dionisio et al., 2013). Therefore, this research on trust toward VHs in VR and the factors that influence trust is essential since trust is a fundamental factor that influences user's engagement, willingness to interact, and acceptance of VHs, thereby shaping the success and effectiveness of virtual environments.

Interpersonal trust encompasses the trust directed towards VHs, including avatars and agents. Avatars can be seen as extensions of users, with social interactions between avatars essentially occurring between the users themselves (Freeman et al., 2020; Freeman & Maloney, 2021; Graber & Graber, 2010). Intelligent virtual agents, on the other hand, are viewed as engaging in a fundamentally social relationship with humans, guided by the same social norms as human-to-human interactions (Nass, Steuer, & Tauber, 1994). Trust, a crucial element in establishing and maintaining social relationships among humans, undeniably holds significant importance in this context.

According to the paradigm, *Computers Are Social Actors* (CASA) (Nass, Steuer, & Tauber, 1994); people follow similar social rules and heuristics when they interact with computers as to when they interact with humans (Nass et al., 1993; Nass & Moon, 2000). In this thesis, we consider computers/VHs as social actors and SIPs since we can interact with them. Distinguishing between trust in humans in the real world and VHs in the virtual world, the concept of trust remains relevant. However, it is vital to recognize the shift in focus toward different types of entities. This shift is crucial in VR trust research, as most studies still heavily rely on the trust literature grounded in traditional human-to-human interactions, despite the unique nature of trust in the virtual environment.

To clarify, trusting VHs refers to an individual's tendency to trust virtual representations or computer-generated characters, such as those encountered in VR or video games (Dwivedi et al., 2022). On the other hand, trusting humans refers to an individual's inclination to trust other human beings in face-to-face or traditional interpersonal interactions (Rotter, 1967). For the avatars, the technical aspects, their technical implementation, and trust in these implementations are part of the trust concept. When interacting with an avatar, users cannot see the ‘real’ person controlling. Thus they cannot verify the identity the avatar claims to be (Lake, 2020). Some individuals may exhibit a higher propensity to trust avatars due to factors such as a sense of anonymity, positive prior experiences with avatars, or a belief in the reliability of virtual representations (Cochard et al., 2004; Machneva et al., 2022; Maloney et al., 2020).

The avatar’s appearance, voice, and other attributes can be altered by its owner, and the verbal or non-verbal communication of avatars is limited due to the current state of VR technology (Radiah et al., 2023). Conversely, others may have a higher propensity to trust humans due to the perceived authenticity, shared cultural norms, or social cues that are absent or limited in avatar interactions (Dwivedi et al., 2022; Oyebode & Nicholls, 2020; Principe & Langlois, 2013). These factors contribute to the differentiation between trust in humans and VHs. This thesis focuses on trust towards VHs in VR, particularly emphasizing the subjective evaluation of the VH’s trustworthiness and the consequential trust-related behaviours given the potential of social VR to connect the physical and virtual worlds. The virtual environments part of the experimental studies in this thesis resembles situations where decisions need to be made and the VHs look like real people in real-life environments.

This thesis builds on previous research that investigated social interaction with VH’s in VR. For example, a review article summarizes research findings from several studies conducted in the field of affective neuroscience, which explored the distinction in evaluation between human-like avatars and actual humans (De Borst & De Gelder, 2015). The objective of the review was to examine how the evaluation of human-like avatars and androids differs from humans in social and affective neuroimaging studies, with a focus on appearance, emotions, action representation, and the potential for VR stimuli in simulating social interactions (De Borst & De Gelder, 2015). During the review, it was pointed out that individuals may experience discomfort when interacting with VHs that closely resemble real humans. This phenomenon is known as the "*uncanny valley effect*," where the discomfort increases as the similarity between VHs, and real humans becomes more pronounced. Interestingly, this discomfort is particularly heightened when the appearance of the VHs is ambiguous, meaning they are not convincingly human-like, compared to when they have a more explicitly human-like appearance, such as in avatars or humans. In other words, when the VHs fall into a middle ground of being almost but not quite realistic, people tend to feel more uncomfortable than when the resemblance is more evident. Furthermore, the review highlighted that human-like avatars could evoke similar emotional responses to humans, albeit with potential differences attributed to physical dissimilarities. The researchers conclude that VR has the potential to replicate complex social situations (De Borst & De Gelder, 2015).

The subsequent experimental studies highlight the importance of considering human-like traits in the VHs included in this thesis. An experimental study investigating a person's response to computers with personality-like characteristics suggests that humans will respond to computers as if they have personalities, regardless of their belief that computers do not have personalities (Nass et al., 1995). A more recent experimental study investigated how trust in embodied virtual agents is influenced by the agent's cooperativeness and an individual's prior experience with other agents (Moradinezhad & Solovey, 2021). The findings showed that participants had higher trust and performance with the cooperative agent, and their trust in the cooperative agent was higher if they had interacted with an uncooperative agent before (Moradinezhad & Solovey, 2021). Furthermore, they confirm that the same factors associated with interpersonal trust in humans apply to interpersonal trust between humans and automated agents (i.e., good intentions, reliable behaviour, assumption of dependability, confidence and competence)(Moradinezhad & Solovey, 2021).

An empirical study on the trustworthiness of life-like interface agents found that patients felt more comfortable sharing personal information specific to their mental and or physical health with a computer-controlled VH (agent) compared with a human-controlled VH (avatar) (Mulken et al., 1999). They demonstrated the potential of computer-controlled VHs to establish a higher level of trust and facilitate the disclosure of sensitive information in healthcare settings. Gombolay et al. (2018), suggest that embodied and anthropomorphic systems are more trustworthy than humans, and as a result, users are more tolerant of errors made by robotic agents than humans. Another study on traditional computer systems reports that user trust was higher in traditional computer systems (windows, icons, menus, pointers) compared with avatars and humans who did the same task (de Visser, 2012). Furthermore, the study uncovered that as the agent's humanness increased, individuals exhibited a higher capacity to maintain trust in the agent even in challenging or uncertain situations. According to de Visser, (2012), establishing trust with a VH may require more time than other entities. However, once trust is established, it becomes more durable. Trust in VHs can be vulnerable to exploitation and misuse, for example, how trust can be manipulated in real-life scenarios. In VR, malicious actors or deceptive entities may deceive users by portraying VHs as trustworthy and reliable, leading users to disclose sensitive information or engage in behaviours they would otherwise avoid. Just as trust can be abused in interpersonal relationships, the misuse of trust in VHs can result in privacy breaches, manipulation, or even harm to users. It highlights the importance of understanding the potential risks and implementing safeguards to protect users' privacy and well-being in virtual environments. These findings emphasize the relevance of trust research in understanding the dynamics of trust towards VHs in VR, as well as the potential risks associated with misplaced trust, for example, in social VR healthcare applications.

VR as a Research Tool

The immersive effect of VR allows social science researchers to gain rich insight into experimental participants behavioural - and subjective experiences (Blascovich et al., 2002), specifically when this data is collected covertly and continuously. By creating virtual environments that simulate real-world scenarios, researchers can observe and analyze how individuals behave, react, and respond within these virtual contexts. The level of immersion offered by VR allows participants to engage with the virtual environment and interact with VH or stimuli in a more natural and intuitive way, closely resembling real-life experiences (Blascovich et al., 2002).

This immersion enables researchers to capture a wide range of behavioural data, such as body movements, gestures, and spatial interactions, which may be difficult to observe in traditional research settings (Bombari et al., 2015; Brookes et al., 2020). Additionally, VR facilitates the collection of subjective data through self-report measures, interviews, and real-time feedback, allowing researchers to explore participants' thoughts, emotions, and perceptions during their virtual experiences. By combining behavioural and subjective data, researchers can gain deeper insights into how individuals navigate and make sense of virtual environments, their decision-making processes, and the underlying psychological mechanisms at play (Wingler et al., 2020).

Examples of constructs that have been studied in VR include trait anxiety, prejudice, trust, racial bias, fear, phobia, and pain (Andreatta et al., 2020; Andreatta & Pauli, 2015; Baker et al., 2020; Bombari et al., 2015; Dunn & Schweitzer, 2005; Lange & Pauli, 2019; Yaremych & Persky, 2019). VR as a research tool has great benefits, as it allows for more experimental control over variables and enables ecologically valid experimental conditions and replication of experimental studies (De La Rosa & Breidt, 2018). A seminal article in 2002 detailed the methods in which VR could provide substantial methodological benefits as a tool for researching social psychology, especially concerning the precise measurement of non-verbal, physical movement behaviour throughout an entire experiment (Blascovich et al., 2002). Therefore, social VR has become an important tool for studying phenomena in social psychology. *Social psychology* is "an attempt to understand and explain how the thoughts, feelings, and behaviour of individuals are influenced by the actual, imagined, or implied presence of others" (Allport, 1985, p 3). VR allows researchers to create virtual environments to study immersed quantifiable experiences, for example, trust in social interaction in ecologically valid settings.

2.3. Measuring Trust toward Virtual Humans

Both subjective and objective measurements have been used for measuring trust toward VHs. Most research on trust towards VH's relied on self-reports as the primary measurement (Liew et al., 2017; Surprenant, 2012) or combined self-reports with other measures such as objective measurements (Aseeri & Interrante, 2021; Hale et al., 2018; Pan & Steed, 2016). The *Trust game* as for objective measurements (Berg et al., 1995) (also known as the investment game) is one of the most popular and established measures of trust in behavioural economics and psychological research. The investment game is played with two players, a trustor and a trustee who are tasked to invest tokens (see section 2.3.1. for a complete description below). The investment game in its original form is played with paper and pencil, as in Berg et al. (1995). However, various experimental studies use a variant of the investment game implemented in VR to study trust behaviour in VR (Abatayo et al., 2020; Atlas & Putterman, 2011; Chiou et al., 2020; Fiedler & Haruvy, 2009; George et al., 2018; Gupta et al., 2020; Lin et al., 2023). For example, Bente et al., (2008) investigate how photorealistic avatars and reputation scores affect trust-building in online transactions using the investment game. Hale et al. (2018) used the investment game to test specific trust towards interactive virtual characters. They found that the results of different characters are highly correlated, which suggests that the investment game measures generalized trust rather than specific trust (Hale et al., 2018). Furthermore, as alternatives, recently, advice-seeking behaviour and the *ask–endorse paradigm* (Clément et al., 2004; Hale et al., 2018; Pan & Steed, 2016) have 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 & Steed (2016), conducted a comparison study of trust among avatar-, video-, and robot-mediated interaction by asking participants to complete a quiz and recording the times they asked and followed advice from two advisors randomly selected from the three alternative representations. Hale et al. (2018), also adopted the *ask-endorse paradigm* to measure specific trust toward virtual characters using a *Virtual Maze* task. Their work has inspired the design of our paradigm in Chapter 4, study 2 or Lin et al. 2023, see section 2.3.2 for a complete description below.

2.3.1. Investment Game

In the *Investment Game* (Berg et al., 1995), the trustor can transfer a certain fraction p of a monetary endowment given to the trustee, while the transferred fraction is increased 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 fraction 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. The experimental studies in this thesis used a VR implementation of one round of the investment game, based on Berg et al. (1995) (see study 2 in Chapter 4, study 3 in Chapter 5, and study 4 in Chapter 6 for the implementation).

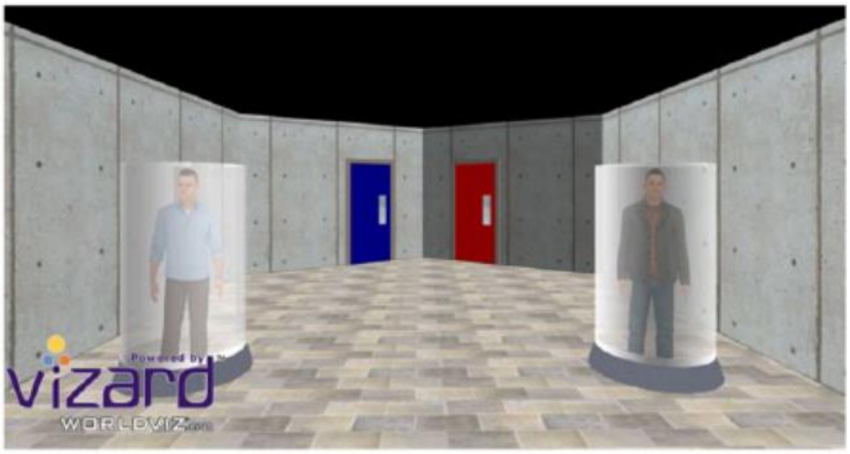
2.3.2. The Virtual Maze paradigm (Hale et al., 2018)

Hale et al., (2018) 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 (Koenig et al., 2004; Koenig & Echols, 2003; Koenig & Harris, 2005, 2007; Pasquini et al., 2007) (see Figure 2). Their work focused on the measurements of generalized trust versus specific trust.

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 that appear 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) (see

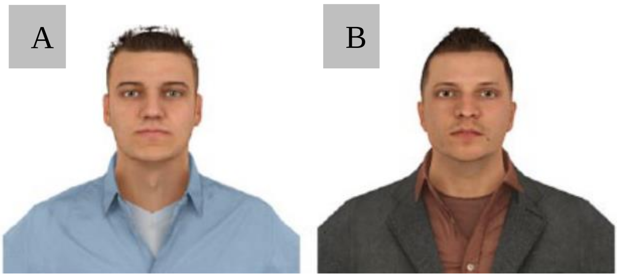
Figure **3**). When the virtual characters are approached, they will suggest a door. The participants keep making decisions until they are notified that they escaped from the maze. Unknown to the participants, there are no right or wrong decisions on the way out of the maze in this task. Rooms and corridors are automatically generated until enough trials (rooms) are observed, and the participant 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 from each character.

Figure 2: Screenshot of Each Room in the Virtual Maze in Hale et al. (2018)



Note: Participants allegedly had to escape from the virtual maze, by entering as few rooms as possible. Two virtual characters were present in a hologram chamber (in each room). Participants had to approach the virtual character(s) to receive navigation advice. After 12 trials participants were told that they had successfully escaped from the virtual maze.

Figure 3: Virtual Characters Present in Hale et al. (2018)



Note: Participants were introduced either during an interview before the virtual maze task. A) virtual character named Mike was designed to appear reliable and provided prompt confident responses, B) virtual character named Ryan was designed to appear unreliable and provided delayed hesitant responses. Both characters were aimed to appear equally likeable and spoke with the same friendliness and clarity in the Hale et al., (2018) study.

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 behaviour 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. The virtual maze allows one to measure trust towards a specific virtual character rather than the general propensity to trust others in general. Thus, it could help measure interpersonal trust towards VHs (see Chapter 4, study 2 or Lin et al. 2023).

2.4. Research Objectives

The previous sections have outlined the critical aspects of trust toward VHs in social VR. Altogether, the above theoretical background shows the complexity of trust research in general, the fundamental characteristics of VR and social VR, the importance of research on trust toward VHs in VR, and the factors influencing it. Furthermore, we demonstrated the lack of a suitable trust measurement tool that explicitly measures trust toward VHs in VR. Therefore, the present dissertation project aims were twofold: 1) to develop and validate a behavioural measurement tool for measuring interpersonal trust toward a VH in VR. 2) To investigate how psychological factors influence trust toward a VH with experimental studies conducted in VR.

As a first step in developing the trust measurement tool, we identified the VR-adjusted version of the social conditioned place preference paradigm (SCPP) (Kiser et al., 2022) as a potential trust measurement tool. We used this paradigm and other self-defined outcome measures to test whether the familiarity of a VH influences trust. The familiarity of a VH was particularly of interest since users in the virtual world can choose to present themselves as "familiar" (e.g., in such a way that others can recognize them: photo-realistic avatar that look just like them in the real world) or "unfamiliar" (e.g., in an unrecognizable character: using a cartoon-like character that hides their true appearance) to those who know them in real life. Furthermore, it is important to know how the familiarity of a VH influences trust, as users trust can be exploited by manipulating users to trust and engage in certain behaviours. For example, placing your trust in a well-known bank advisor (whose identity has been compromised) within a VR setting and subsequently having your financial information exploited by the cybercriminal, leading to the depletion of funds from your bank account. In this between-subject *familiarity stud*y (in Chapter 3), half of the participants were familiarized with one VH, and the other half with another VH and they were tasked to interact with the VHs in a virtual art museum. The goal being to determine if participants evaluate the familiar VH as more trustworthy than the unfamiliar VH? Additionally, aside from participants subjective evaluations of VH trustworthiness, will their behaviour align with their trust assessments of the VH? For example, will participants stand closer, share more information, respond quicker to the familiar VH, and spend more time in the room with the familiar VH? Understanding how familiarity with a SIP influence trust can assist with the development of identity management systems in virtual social environments where social interaction through VHs is possible, for example, transactions in VR.

The results of the *familiarity stud*y (in Chapter 3) and the limitations of the virtual maze paradigm of Hale et al. (2018) was considered in the construction and validation of our trust measurement tool, in the between-subject design in study 2 (*validation study*) (in Chapter 4). Participants interacted with a trustworthy or untrustworthy VH in a virtual maze of 12 rooms. Participants were tasked to escape from the virtual maze by entering as few rooms as possible. The VH that was present provided navigation advice on request, and participants had a choice to follow the advice. A real human allegedly controlled the VH and supposedly had a map of the maze and knew the way out. The subjective evaluations of the VH's trustworthiness and the consequential trust behaviour (asking and following advice and the time to execute the advice) was measured. It is important that our paradigm is sensitive to behavioural differences in the VH's trustworthiness between the two conditions. We investigated whether participants in the trustworthy condition evaluated the VH as more trustworthy and, as a result, showed more trust behaviours, for example, asking and following more advice. This validated behavioural tool is a novel contribution to the academic community researching trust toward VHs in VR and was used in the following studies.

Based on the results of the study 2 (as described in Chapter 4), the paradigm was used to investigate the second research aim, how psychological factors influence trust toward VHs, with two within-subject experimental studies conducted in VR. Study 3 (as described in Chapter 5) investigated the influence of cognitive load (high and low cognitive load conditions) on trust toward VHs, while study 4 (as described in Chapter 6) examined the impact of emotional affect (negative and positive emotional affect conditions) on trust toward VHs. We were interested in observing whether cognitive load (in study 3, in Chapter 5) and emotional affect (study 4, in Chapter 6) influenced participants' evaluation of a VH's trustworthiness. Furthermore, do participants ask and follow advice more or less, and do participants execute the received advice faster or slower between the conditions of the two studies? Researching the relationship between cognitive load, emotional affect, and trust toward VHs in VR yields valuable insights that can improve learning environments, enhance user experiences, and ensure user safety and security in VR settings.

In social VR, users engage in various tasks, including gaming, education, virtual tours, and social interactions. Some of these tasks can be mentally demanding and require significant cognitive resources. By understanding how cognitive load affects trust in VHs, researchers can enhance the design of educational and interactive experiences, ensuring optimal learning environments and improved user performance. Emotional states greatly influence users' behaviour and overall experience in the virtual world. Investigating how emotional affect impacts trust in VHs enables developers to create immersive environments that elicit positive emotions, foster trust, and enhance user engagement and satisfaction. Furthermore, given the potential for malicious actors to manipulate trust in virtual environments, exploring the interplay between cognitive load, emotional affect, and trust becomes crucial. By understanding how scammers exploit cognitive and emotional factors to deceive users and influence their behaviour, researchers can develop effective countermeasures, security protocols, and user awareness strategies to safeguard individuals from potential harm, such as identity theft, fraud, or other malicious activities.

3. Study 1: Who you meet

Familiarity and Trustworthiness in Virtual Social Interaction Partners

3.1. Introduction

People prefer, approve, and trust entities that they are familiar with (Gulati & Sytch, 2008; Uzzi & Gillespie, 2002; Vugt et al., 2010). Therefore, the familiarity of a VH may influence the evaluation of their trustworthiness. We identified proxies of trust that can be used to measure trust based on past literature, which includes the subjective evaluation of a social interaction partner (SIP), interpersonal distance, disclosure of information, and response time. The familiarity of a virtual avatar leads to a sense of closeness, similarity, liking, and a feeling of intimacy or closeness (Jamieson, 2007; Park et al., 2021). Experimental studies found that repetitive exposure to a face led participants to describe it as more familiar, similar, and attractive (Koca & Oriet, 2023; Peskin & Newell, 2004). Perceptions of familiarity are related to trust judgments, especially when social reality is shared (Echterhoff et al., 2005). An experimental study investigated how mere exposure influences personal preferences and trust (Kwan et al., 2015). Participants had to assess the trustworthiness of a company's employees after subliminal exposure (either 18 or 2 times) to the employees. Participants read the employee profiles, which were similar in content and randomly paired with the names of the employees (Kwan et al., 2015). The study found that frequent exposure increases perceived familiarity, acceptance, and trust (Kwan et al., 2015). Previous research found that prolonged social interaction increases likability towards SIPs (Mackenzie, 1948; Wilner et al., 1952). It was later suggested by Zajonc (1968), that mere exposure to a stimulus (including a SIP) leads to higher interpersonal attractiveness. Thus, the more a person is exposed to something or someone, the more familiar and comfortable they become with them (Koca & Oriet, 2023; Peskin & Newell, 2004). This familiarity perception may lead to trust and positive responses (Kwan et al., 2015). For example, people approach and keep close proximity to familiar and pleasant entities and avoid and increase proximity to unfamiliar and unpleasant ones (Hall & Hall, 1966; Vugt et al., 2010).

Vugt et al., (2010) investigated how embodied agents' facial similarity influence users' responses towards virtual agents. Specifically, psychological involvement (approach tendencies like empathy or sympathy), psychological distance (avoidance tendencies like irritation or boredom), and the intention to use the virtual agent in the future. During the experiment, participants had to complete a science trivia task where an agent was present to provide advice of which the helpfulness varied between tasks. The agent was designed to either look similar or dissimilar to the user. The gender of the agent matched the gender of the participant. The experiment relied on self-reported measures, examples include the following questions, “How much does X appeal to you?”, “How connected do you feel to X?” and, “How much do you want to use X again?”. The results revealed that facial similarity had increased participants' ratings of intention to use the agent (in females) and involvement (in males and females). Male participants had a negative response to an unhelpful similar-looking agent compared to the dissimilar one. Furthermore, the study found that participants feel psychologically closer to familiar and pleasant agents and avoid unfamiliar and unpleasant ones (Vugt et al., 2010).

Building on this, work Higgins et al. (2022) showed that participants who perceive the VHs as unpleasant feel uncomfortable with them. In two experimental studies in VR Higgins et al. (2022) investigated whether the “mismatch between the synthetic voice and the photorealistic appearance of the character would reduce the comfort with the character, lower social presence and decrease appeal, familiarity and increase eeriness” (Higgins et al. 2022, p.117). Physical distance (proximity) was used as a measure of comfort, with the underlying assumption that people stand closer to familiar or pleasant people and further away from unfamiliar people (both real life and in VR) (Bailenson et al., 2003; Hall & Hall, 1966; Higgins et al., 2022). Higgins et al. (2022) placed the participant intentionally closer than 40 cm to the VH in the virtual room to investigate if they felt uncomfortable with the close proximity to it (40 cm is the personal space distance participants usually keep, as per Bailenson et al., (2003) in VR). To measure how comfortable participants were in different emotional scenarios (sad, friendly, and unfriendly), they were asked to answer “yes” or “no” to the following question: “When I first saw the girl in the room, I felt I was standing too close, I was in her intimate space”. The analysis of proximity revealed differences according to the emotional scenarios. Participants reported higher discomfort with the closeness of the character in the sad condition and more comfort in the friendly condition. However, no significant differences in the pairwise comparisons were observed (Higgins et al. 2022).

Furthermore, an experiment investigating cooperation during a once-off encounter under uncertainty found that the more familiar participants are with the SIP, the more personal information they disclose (Wheeless & Grotz, 1977). In any interpersonal relationship, the individuals involved develop a level of familiarity with one another over time through repeated interactions and shared experiences (Cochard et al., 2004; Rosenberger et al., 2020). As a result, they become more knowledgeable about each other's character, values, and behaviours, which allows them to develop a sense of predictability and reliability about the other person

(Koenig & Harris, 2007; Rosenberger et al., 2020). Furthermore, the specific SIP also plays a critical role in the trust-building process because different individuals have varying degrees of trustworthiness, depending on their past behaviour, reputation, and other characteristics (Johnson-George & Swap, 1982). Thus, people are more likely to trust someone who has demonstrated trustworthiness in the past, while they may be less likely to trust someone with a history of untrustworthy behaviour. The nature of the interpersonal relationship and the specific SIP are essential to trust in any dyadic relationship because they build familiarity, share experiences, and mutual understanding, which are critical components of trust. Moreover, rapidly formed automatic appearance-based impressions strongly influence the expectation that a SIP will cooperate (Chang et al., 2010; Engell et al., 2007; Olivola & Todorov, 2010; Rezlescu et al., 2012). These impressions guide people's trust behaviours in social situations, for example, whether to share or withhold information, cooperate or not to cooperate, follow advice, etc. People tend to over-rely on appearance-based judgments as they make inferences about another person's character, which can be completely incorrect (Olivola & Todorov, 2010). Familiar people are not always trustworthy, and unfamiliar people are not always untrustworthy. However, a set of experiments has demonstrated that familiarity contributes to the automatic formulation of impressions. Although it takes 100-ms exposure to a facial stimulus to form an impression about a depicted person, it is not enough time for information to be processed consciously (Willis & Todorov, 2006). Interestingly, extended exposure to the individual's face does not substantially modify these initial impressions, although it may enhance one's confidence in the accuracy of their judgments (Wargo, 2006; Willis & Todorov, 2006). Familiarity contributes to the automatic formulation of impressions (Bonnefon et al., 2013; Olivola & Todorov, 2010). Based on the reviewed literature and the influence familiarity has on trust, we experimented to investigate how the familiarity of a VH influences a participant's trust evaluation of the specific VH and the corresponding trust behaviours. In this study, we aimed to answer the research question: does familiarity with a SIP influence trust evaluation and behaviour?

From the experimental studies reviewed in Chapter 2, we conclude that no subjective or behavioural measure is suitable for measuring interpersonal trust toward VH's in VR. Although repetitive exposure to faces gained researchers attention in the past (Koca & Oriet, 2023 & Peskin & Newell, 2004), how familiarity with a VH influences trust was not investigated in VR before. Thus this study is the first to investigate whether the familiarity of a VH influences interpersonal trust in VR, with a social conditioned place preference paradigm (SCPP; (Kiser et al., 2022). Animal researchers' developed the SCPP to investigate social approach-avoidance mechanisms in rodents (Calcagnetti & Schechter, 1992). Conditioning paradigms are standard in animal research because they provide a controlled and standardized way to investigate the underlying mechanisms of animal learning and behaviour, which can then be translated into humans (Calcagnetti & Schechter, 1992). Kiser et al. (2022) adjusted the SCPP paradigm to study human behaviour in VR objectively. The following familiarity study explored the impact of virtual human familiarity on trust. The SCPP paradigm (Kiser et al., 2022) (paradigm is explained below) and other literature-inspired outcome measures as mentioned above. These include proximity to the virtual human, amount of information shared with the VH, response time, and the place preference effect: dwell time, as a potential behavioural paradigm that was used to measure trust towards VH in VR.

The Social Conditioned Place Preference (SCPP) Kiser et al., (2022)

The Conditioned Place Preference (CPP) paradigm is a behavioural paradigm used to study the rewarding or aversive effects of drugs, environmental stimuli, or experiences in animals (Kiser et al., 2022). It involves training animals to associate a particular environment or location with a rewarding or aversive stimulus and then measuring their preference for that environment or location in subsequent tests (McKendrick & Graziane, 2020). CPP is commonly used in preclinical research to study the neurobiological mechanisms underlying drug addiction and to evaluate the efficacy of potential treatments for addiction. The *social conditioned place preference paradigm (SCPP)* is a modification of the CPP that evaluates the rewarding or aversive effects of social cues, including play behaviour, sex, and general social interactions, alongside the effects of substances (Calcagnetti & Schechter, 1992). The SCPP paradigm was adapted from animal research to draw conclusions about the fundamentals of human behaviour and disorders characterized by motivational dysfunctions. An experimental study tested an SCPP paradigm to measure social and emotional processing in children with language impairments. The study results demonstrated that children spent significantly more time in the room with a social interaction stimulus (rewarding adult) following training (Baron et al., 2020). Other SCPP experiments include studying associated learning frequency (Reeb-Sutherland et al., 2011) and designing and constructing a child-friendly space (Hiller et al., 2015), to name a few. Kiser et al. (2022) implemented an SCPP paradigm for humans and investigated the impact of trait social anxiety on the approach and avoidance of VHs in VR. In their study, the virtual environment resembled a virtual art museum with a lobby and two exhibition rooms. Participants were exposed to two virtual agents (computer-controlled) that were either happy or angry in appearance in a specific room. Kiser et al. (2022) explored how social interaction with happy/angry-looking virtual agents influenced participants' dwell time and their subsequent evaluation of the specific room in a subsequent test conducted without the presence of an agent.

The paradigm consists of three experimental phases (see Figure 4):

1) *Habituation phase*– the participants started in the lobby of the virtual art museum; they were instructed via pre-recorded audio to enter one of two rooms (1 or 2). The participant freely explored rooms 1 and 2 for 2 minutes without the possibility of returning to the lobby (the door closed after entering the rooms). After 2 minutes, the participant was teleported back to the lobby. There were no VHs present during the habituation phase. Each participant's position and dwell time (duration of stay) in each room were recorded during this phase.
2) During the *conditioning phase* – participants could again enter both rooms from the lobby. This time the VH (either with a smiling expression for the place preference group or an angry facial expression for the place aversion group) was present in one of the two rooms (rooms were counterbalanced to control room effects). In the other room, a neutral object (computer screen or a desk) was present. Participants were instructed (via an audio recording over the CAVE speakers) to enter a room (e.g., room 1) and look at the paintings on the walls for 1 minute, then verbally report how much they like each image, the agent, the neutral object, the floor and the room itself (the latter was only considered as a variable of interest) on a 10- point Likert scale (0 indicating "Did not like XY at all" and 10 indicating "Liked the XY very much"). Hereafter, participants were teleported back to the starting position (the lobby) and entered the remaining room (e.g., room 2). The procedure was repeated three times in both rooms; the virtual agent and neutral objects were placed at a new location within the same room. After this phase, the participant was teleported back to the lobby.
3) *Test phase* – starting in the lobby, the participant was instructed via a pre-recorded audio to freely explore one of the two rooms for 2 minutes without any VHs (social stimulus) and neutral object present. The position and dwell time (duration of stay) were recorded. The primary outcome was the dwell time.

The SCPP paradigm was identified as a potential trust measurement tool and was used along with other self-defined outcome measures to test whether the familiarity of a VH influenced trust.

Figure 4: SCPP Experimental Procedure of Kiser et al. (2022)

Virtual Reality			
Training 2-4 minutes	**Experiment**		
	Habituation 2 minutes	**Conditioning 6 x1 minute**	**Test 2 minute**

Note: SCPP = Social Conditioned Place Preference; summary of the experimental procedure. During the habituation and test phases, participants freely explored both rooms; the door between rooms 1 and 2 was open. The conditioning phase started with the participants in the lobby; from there, they explored each of the two museum rooms individually six times for 1 minute each (the door between rooms 1 and 2 was closed) (Kiser et al., 2022).

3.1.1. Hypotheses

The reviewed literature inspired the following hypotheses:

H 1: The familiar VH is evaluated as more human and more trustworthy.

H 2: Participants stand closer to the familiar VH.

H 3: Participants share more information with the familiar VH.

H 4: Participants respond quicker to the familiar VH.

H 5: There is a place preference effect during the test phase, participants spend more time in the room in which the familiar VH was present during the previous conditioning phase.

3.2. Methods and Materials

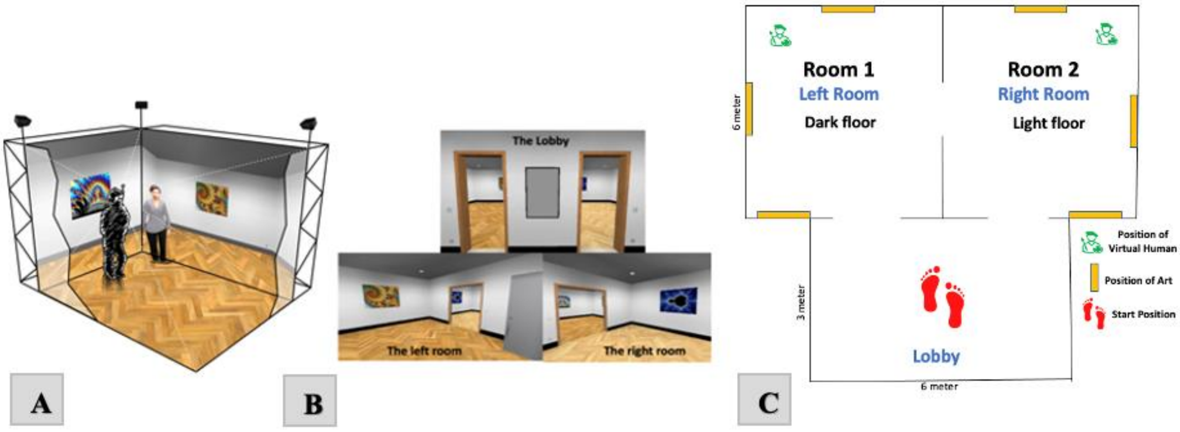
3.2.1. Study Design and Participants

In a within-subjects design, 20 women (*M* = 24.1 years, *SD* = 3.09, range 18 to 35) participated in the experiment. Virtual human familiarity was the within-subject factor, all participants were exposed to a familiar and an unfamiliar VH. The Ethical Committee of the Psychological Institute of the Faculty of Human Sciences of the University of Würzburg reviewed the protocol and provided ethical clearance (GZEK 2020-97).

The inclusion criteria consisted of female participants with no psychiatric or neurological disorders. To match the gender of the participants, the virtual humans were also female. Participants were fluent in German and healthy; those who indicated claustrophobia in the recruitment phase were excluded from the study. 86% of the participants participated in a VR experiment before. The recruitment of participants relied on the university participant database (SONA), and participants received 12 euros for their participation in the study, which lasted one hour. Participants were randomly assigned to the conditions and the order of the conditions was counterbalanced.

3.2.2. Apparatus

Figure 5: The Virtual Environment in the Computer Augment Virtual Environment (CAVE)

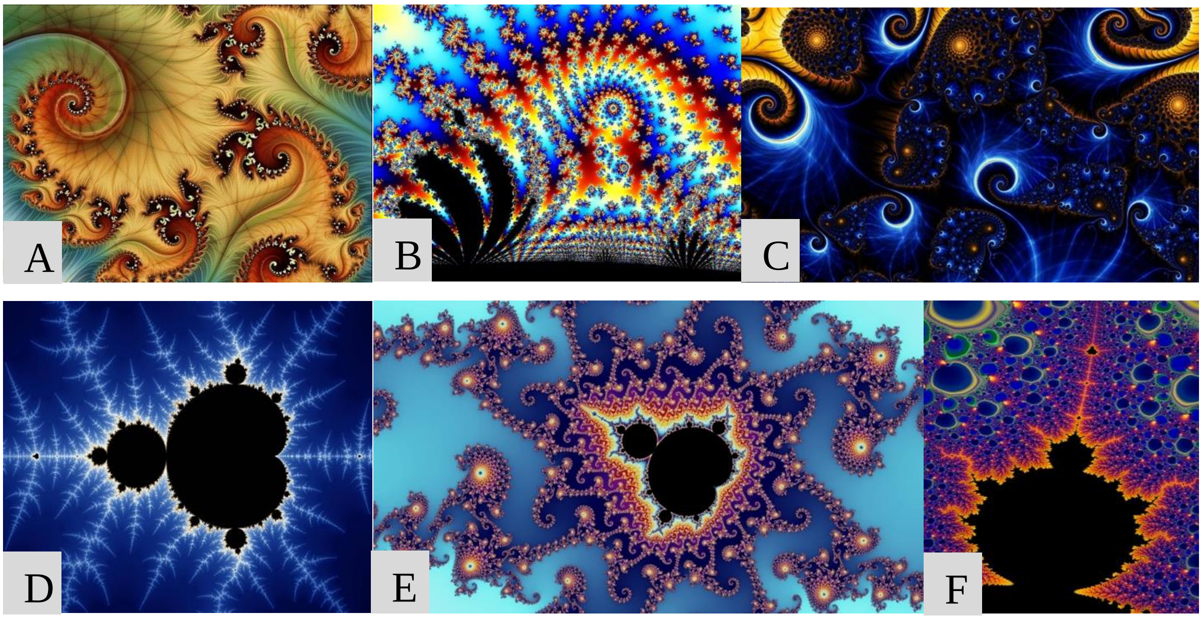


Note: A) Schematic representation of a participant in the 5-sided Computer Augmented Virtual Environment (CAVE) and a VH. B) A screenshot of the virtual environment showing the different rooms stereoscopically presented in 360° and the floor. The lobby room was always the starting position of the participants. C) Exemplary sketch of the virtual environment; yellow blocks are the art positioning, and the green figures in the upper corners of the two rectangular rooms indicate the position of the VHs.

The participants wore interference-filtering glasses for 3D effects (Infitec Premium, Infitec, Ulm, Germany). They utilized a gamepad (Xbox 360 Wireless Controller, Microsoft, Redmond, WA, USA), which allowed them to move freely in the confines of the five-sided CAVE at the University of Würzburg (previously described by our department in Gromer et al., 2018). The participant's movement and orientation were tracked with an infrared LED system using four cameras positioned in the four corners of the CAVE (PhaseSpace Impulse, PhaseSpace Inc., San Leandro, CA, USA). The five sides of the CAVE are 4 x 3 x 2.95 meters, on which the virtual environment was projected with six stereoscopic image projectors (Barco GALAXY NW7). The projected images on the floor and door wall had a resolution of 1627 x 1200 pixels, the resolution on the front wall was 2016 x 1486 pixels, and the two smaller side walls had 1220 x 1200 pixels (Kiser et al., 2022). The sounds for the instructions were provided with a 7.1 surround sound system.

The *virtual environment* was rendered with a modification based on the Source Engine SDK 2007: VrSessionMod 0.5 (a) (Valve, Bellevue, WA, USA). To script the event procedures during the experiment and data acquisition, we used: CS-Research 5.6 software (VTplus, Würzburg, Germany; see www.cybersession.info for detailed information). The virtual experimental environment consisted of two (6 x 6 m) rectangular rooms and an adjacent lobby (3 x 6 m) (see Figure 5) (based on Kiser et al., 2022). The participants started the experiment in the lobby, and they could enter two rooms (all rooms were connected through doors). To ensure differentiation between rooms, the floor colour (dark or light brown) and the images in the two rectangular rooms (three neutral Mandelbrot images in each room, see Figure 5 and Figure 6) were consistent throughout the experiment. The virtual environment resembled a virtual museum with two rooms distinguished from each other by the floors and the pictures on the walls of the rooms. This design instinctively induced exploration (Kiser et al., 2022).

Figure 6: Mandelbrot Images



Note: Images A, B and C used in room 1 and images D, E and F used in room 2

Virtual humans. The VHs were the same as those in the study of Kiser et al. (2022), and the familiar and unfamiliar VH were counterbalanced between rooms. Faceposer from the Source Engine SDK toolkit (Valve, Bellevue, WA, United States) was used to construct the facial expression animations (based on the faces of real humans) for the two VHs used in this experiment (see Figure 7). The VHs had slow body movements (based on the idle posture of Half-Life 2) and continuous scripted facial expressions (lasting 60 seconds). The computer-controlled VHs could not directly interact with the participants, but their gaze followed the position of the participants.

Figure 7: Virtual Humans

Note: Screenshots of the virtual humans used in the Social Conditioned Place Preference Paradigm. Half (n=10) of the participants were familiarized with VH A, and the other half (n=10) were familiarized with VH B. VH C was present only in the tutorial room.

3.2.3. Familiarity Manipulation and Measurements

Manipulation of Familiarity

Before starting the SCPP paradigm, participants took part in a tutorial session where they learned how to navigate in the CAVE; by traversing a virtual maze. After that, participants entered the lobby of the virtual art exhibition, where the familiarity manipulation of one VH (see Figure 7 either A or B) took place. In the lobby, participants looked at a photo-realistic picture of the VH for 2 minutes while a pre-recorded audio was played. The pre-recorded audio consisted of information on how the VH was allegedly created based on the artist's appearance in real life. Personal information of the artist in real life was shared, such as their name, relationship status, and years of experience (Appendix B). To avoid order effects, one familiar VH and one unfamiliar VH were randomly assigned to participants. The familiar VH was positioned either in room 1 or 2, and the positions were counterbalanced across participants during the experiment.

Questionnaires

During the *pre-VR measures*, participants reported demographic information, health questions, Big-Five Inventory (BFI; Rammstedt & John, 2007), STAI state, and trait (Laux, 1981). The BFI is an abbreviated 10-item version of the 44-item BFI and was used in this study to measure participants' personality traits (Rammstedt & John, 2007). The subscales include extraversion, agreeableness, conscientiousness, neuroticism, and openness, and a mean score between 1 and 5 is calculated for each subscale. In addition, participants had to rate on a 5-point scale how well the items describe their personality. The scale ranged from 1 "Disagree strongly" to 5 "Agree strongly", resulting in a mean score between 1-5 for each dimension. Examples of items include: "I see myself as someone who is reserved", "I see myself as someone who is generally trusting", or "I see myself as someone who tends to be lazy" (Rammstedt & John, 2007). The STAI is a self-report questionnaire that measures anxiety as a state and trait (Laux, 1981). For the state anxiety subscale, participants are asked to rate 20 items (e.g., "I feel frightened") according to their present feelings on a four-point Likert scale ranging from "not at all" to "very much so." For the trait anxiety subscale, participants are asked to rate 20 items (e.g., "I feel nervous and restless") according to how they feel in general, on a four-point Likert scale ranging from "almost never" to "almost always." A sum score with a range of 20–80 is calculated for each subscale. A German version of the STAI scales was used (Laux, 1981).

The *post-VR measures* included the SSQ (Kennedy et al., 1993) and the IPQ (Schubert, 2003). The SSQ is a self-report questionnaire that measures simulator sickness, i.e., symptoms such as nausea, dizziness, headache, or eyestrain, caused by immersion into virtual environments (Kennedy et al., 1993). The questionnaire consists of 16 items rated on a 4-point Likert scale ranging from "none" to "severe." The resulting weighted mean scores represent the three factors nausea (e.g., stomach awareness), oculomotor problems (e.g., eyestrain), and disorientation (e.g., vertigo), and a total score (Kennedy et al., 1993). The IPQ is a self-report questionnaire that measures presence, the sense of 'being there' in virtual environments. The questionnaire consists of 14 items rated on a seven-point Likert scale. The IPQ measures three subscales representing different dimensions of presence. The spatial presence subscale measures a feeling of being inside the virtual environment (e.g., "Somehow, I felt that the virtual world surrounded me"). The involvement subscale consists of items measuring an attentional focus toward the virtual environment (e.g., "I was not aware of my real environment"). The experienced realism subscale measures how real the virtual environment seems to the participant (e.g., "How real did the virtual world seem to you?"). One additional item measures a general sense of being in the virtual environment ("In the computer-generated world, I had a sense of 'being there"). Each item of the IPQ is scored on a 7- point scale ranging from 0 "Totally disagree" to 6 "Totally Agree". Table 2 summarizes the descriptive questionnaire data per group.

Table 2: Descriptive Questionnaire Data

	Familair with VH **A**		Familiar with VH **B**	
	M	**SD**	**M**	**SD**
Pre – Questionnaires				
Age	24.17	3.26	24.50	3.00
BFI Extraversion	3.33	1.11	3.47	1.02
BFI Agreeableness	3.22	0.67	3.17	0.64
BFI Conscientiousness	3.75	0.77	3.78	0.73
BFI Neuroticism	2.64	0.54	2.75	0.60
BFI Openness	3.94	1.06	4.08	1.02
STAI Trait	44.00	11.14	43.00	11.00
STAI State	42	7.45	41	6.42
Post – Questionnaires				
SSQ Total	23.06	23.28	19.95	22.22
SSQ Nausea	15.90	23.14	14.31	22.49
SSQ Oculomotor	16.42	15.87	13.48	14.60
SSQ Disorientation	32.48	36.98	28.61	37.13
IPQ Spatial Presence	4	0.6	4	1
IPQ Involvement	4	0.9	4	1
IPQ Experienced Realism	3	1	3	1
IPQ General	4	0.5	4	1

Note: BFI = Big Five Inventory; STAI = State-Trait Anxiety Inventory; SSQ = Simulator Sickness Questionnaire; IPQ = Igroup Presence Questionnaire.

Main Outcome Measures

The CAVE system recorded the movement and orientation of the participants. The experiment included three phases (habituation, conditioning, and test phases), analogous to the work of Kiser et al. (2022). In the *habituation* and *test phase,* no VHs were present, participants were tasked to freely explore the two rooms for 2 minutes. The main outcome measure the place preference effect which was calculated as follows:

The time (in seconds) spent within the room that was previously associated (during the *conditioning phase*) with the social cue (familiar VH) was recorded. Given that both the *habituation* and *test phases* had a duration of 2 minutes each, the difference between the time spent in this room during the test and habituation phases indicated a change in place preference.

In the *conditioning phase,* participants interacted with a familiar VH in one room and an unfamiliar VH in the other room. The following behavioural proxies of trust were measured and compared between the two rooms in each experimental phase:

1) the time spent in each room measured in seconds and milliseconds,

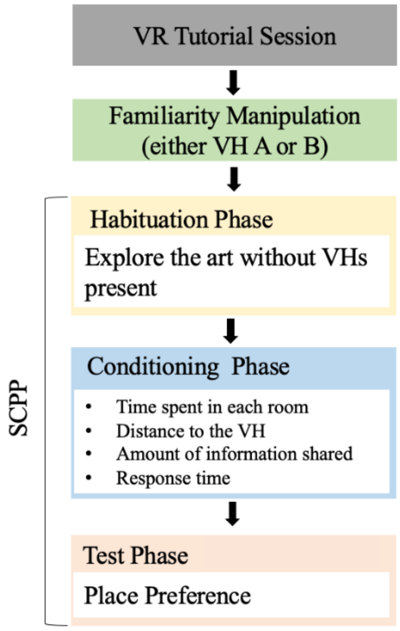
2) the participant's distance to the VHs during the first approach task measured in meters,

3) the amount of information shared with the VH, measured by counting the number of words shared,

4) the response time was measured, by measuring the silence between the VH's question and the participant's answer in seconds.

3.2.4. Procedure

Figure 8: Summary of the Experimental Procedure



Note: SCPP= Social Conditioned Place Preference, VR = Virtual Reality, VH = Virtual Human.

At the beginning of the study, participants read the study information and provided informed consent, then completed the *pre-VR measures*. Participants were informed that they were participating in a study where they looked at art in more than one room in a virtual museum where VHs would be present (see Figure 7A or B). Before participants entered the CAVE, they were equipped with a microphone and 3D glasses. Then they completed a short maze-like *tutorial* of 2 - 4 minutes where they learned how to use the game controller to navigate by practicing approaching VH C (see Figure 7C). The experimenter used a microphone to communicate with the participant; all audio feedback was provided via the speakers of the CAVE. After the tutorial session, participants were teleported to the lobby of the virtual museum, where they were placed equidistantly from two doors through which they entered the virtual museum (starting position). When participants were in any of the rooms, they could freely move using the controller to reach other locations in the CAVE. Next, participants were familiarized with one VH (A or B) through the procedure described above in section 3.2.3. (See Figure 7). Participants were then teleported to the lobby of the virtual museum to start the habituation, conditioning, and test phase. After that, participants completed the post-VR questionnaires, followed by a full disclosure of the true purpose of the study. The experimental procedure is summarized in Figure 8.

During the *habituation phase,* participants were instructed to enter the virtual art museum by choosing which room to enter first. After that, they could move freely between both rooms for 2 minutes, after which they were automatically teleported to the starting position in the lobby (see Figure 5). Then followed the 15-minute *conditioning phase*, with the participant in the starting position, they had to enter both rooms; in one room was a familiar VH, and in the other room was an unfamiliar VH (the position of the VH was randomized and counterbalanced). Upon entering a new room, the VH would welcome the participant. This phase had three tasks and two trials (once with the familiar VH in one room and once with the unfamiliar VH in the other room), and participants always entered the left room first:

During the *first task***,** participants were tasked to enter both rooms, look at the art on the walls, approach the VHs (see Figure 7A and B), and ask each of them whether the art is an original art piece (the time spent in each room with the familiar VH and the unfamiliar VH was measured, as well as the distance to the VHs). The VHs would either respond “yes” or “no.” The responses were randomized and counterbalanced. This marked the end of the first task, and the participant was teleported back to the lobby.

In the *second task*, from the starting position in the lobby of the virtual museum, participants had to enter the left room with an open door between the rooms, approach each VH in each room, and tell them that they liked their art (the distance to the familiar and the unfamiliar VH was measured). The VHs responded (pre-recorded audio), “What do you like about my art”?

The *third task* of the participant was to answer the question of each VH, and their audio responses were recorded (the response time to the VH’s question was measured as well as the amount of information shared was measured). This marked the end of the conditioning phase, and the participant was teleported back to the starting position.

During the *Test phase*, participants were asked to enter rooms 1 or 2 and explore the room just like in the pre-conditioning phase for 2 minutes without any VHs present. (The duration of their stay in each room was recorded). After the experiment, participants answered the post-VR measures, and the study's main aim was disclosed in the debriefing.

3.2.5. Statistical Analysis

Jamovi version 2.3.21.0 (The Jamovi project, 2021) was used to conduct *t*-tests for hypothesis testing. First, a one sample *t* – test was conducted to test differences in the familiar and unfamiliar VH's trustworthiness and humanness scores. Then a paired sample *t*-test was conducted to test subjective differences in the VH's trustworthiness and humanness scores dependent on their familiarity. To explore differences between the behavioural outcomes, a series of paired sample *t*-tests were conducted. Finally, the place preference effect was calculated to test whether participants spend more time in the room in which the familiar VH was present during the previous conditioning phase, as explained above. A one samples *t*-test was conducted to test the significance of the difference in the time spent in the habituation and test phase (social place preference effect), in the room previously associated with the familiar VH. The level of significance was set at $p < .05$.

3.3. Results

Subjective Measurements

Table 3: Virtual Human Trustworthiness and Humanness Scores

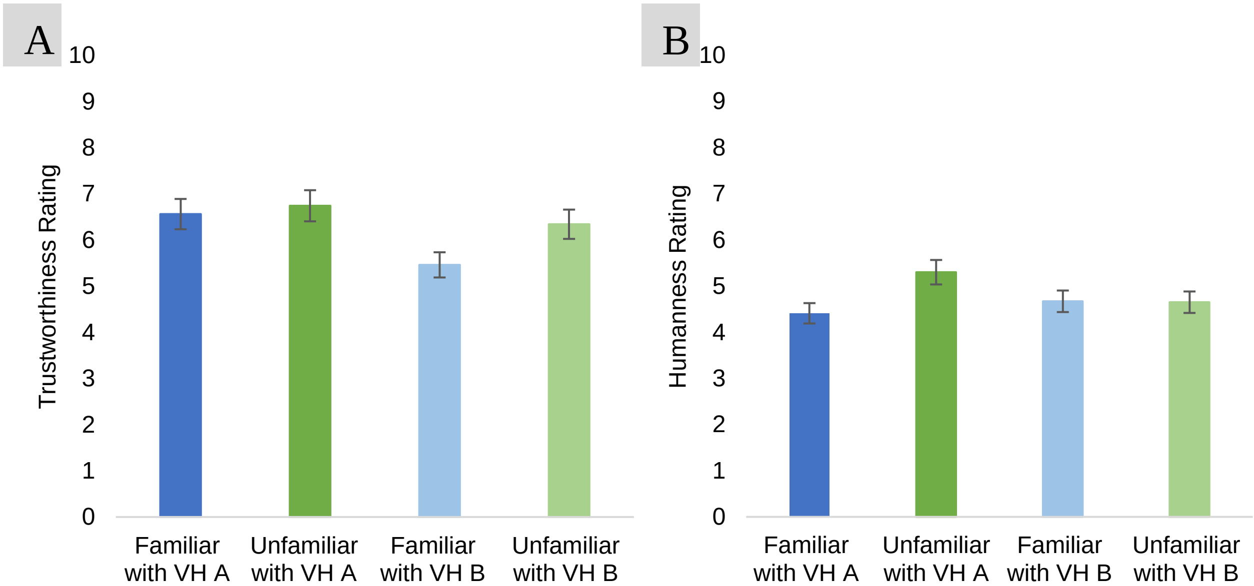
	Familair VH		Unfamiliar VH						
	M	**SD**	**M**	**SD**	***df***	**t**	***p***	***d***	**95% CI**
Familiar with VH **A** Trust score	6.56	1.90	6.34	1.76	9	.60	.566	.19	-.615, 1.06
Familiar with VH **B** Trust score	5.46	2.26	6.74	1.77	9	-1.83	.101	-.58	-2.86, .303
Familiar with VH **A** Humanness	4.41	1.23	4.65	1.93	9	-.69	.510	-.22	-1.03, .552
Familiar with VH **B** Humanness	4.67	2.47	5.30	2.32	9	-.73	.490	.23	-2.61, 1.35

Note: VH = Virtual Human. The trustworthiness and humanness of the VHs were assessed with an explicit subjective question on a 10-point Likert Scale, *** $p < .001$.

A one-sample t-test was initially conducted to assess the mean difference between familiar and unfamiliar VH's trustworthiness and humanness scores. However, the assumption of normality was violated based on the Shapiro-Wilk test (trustworthiness: p = .060, humanness: p = .005). Subsequently, non-parametric Wilcoxon rank tests were performed, revealing no significant differences in trustworthiness (p = .873, one-tailed) and humanness (p = .659, one-tailed) scores between the familiar and unfamiliar VH's.

Results of the conducted t-tests showed no significant difference between the trustworthiness and humanness ratings for the familiar and unfamiliar VH (see Table 3 for a summary and Figure 9 for a visualization).

Figure 9: Virtual Human Trustworthiness and Humanness Scores



Note: VH = Virtual Humans, A) Depicts the trustworthiness scores of the VH's, and B) depict the humanness scores of the VH's. The blue bars indicate the average scores of the familiar VH, and the green bars indicate the average scores of the unfamiliar VH. The error bars indicate the standard error, *** p < .001.

Behavioural Measurements

Table 4: Summary of Behavioural Measurements

	Familiar VH		**Unfamiliar VH**						
	M	**SD**	**M**	**SD**	***df***	**t**	***p***	***d***	**95% CI**
Interpersonal distance Task 1:	1.35	.26	1.55	1.74	19	-.88	.388	-.20	-.063, .248
Interpersonal distance Task 2:	2.11	1	2.22	.93	19	-.28	.782	-.06	-.501,.377
Information shared	12.05	7.43	11.50	5.59	19	.60	.555	.13	-.308, .573
Response time	2.36	.88	2.43	1.74	19	-18	.857	-.04	-.479, .398

Note: Interpersonal distance was measured in meters, information shared was measured by the number of words participants shared and response time was measured in seconds, *** p < .001.

Regarding the behavioural outcomes, there was no statistical difference between participants' interpersonal distance, the amount of information shared, and the response time to the VH when they were *familiar* compared with when they were *unfamiliar* with the VH (see Table 4 for a summary of the statistical tests).

No *social place preference effect* was observed between the time spent in the room that was previously associated (during the *conditioning phase*) with the social cue (familiar VH). The results are summarized in Table 5. Although participants spent more time in the room that was previously associated with the familiar VH, a one sample *t*-test shows the difference in time spent between the habituation and test phase was not significant (p = .564).

Table 5: Social Place Preference Effect (room with familiar VH)

	M	SD
Habituation Phase	56.61	24.75
Test Phase	61.76	24.18
Social Place Preference	5.16	39.31

Note: The time spent (in seconds) in the room that was previously associated with the familiar VH in the *conditioning phase* are reported in both the habituation and test phase. SPP; social place preference effect is the difference in time spent in the habituation phase and test phase.

3.4. Discussion

This study used a potential behavioural paradigm to measure interpersonal trust towards a familiar virtual human. Before participants started with the experimental tasks, they were familiarized with one of two virtual humans. The experiment used the social conditioned place preference paradigm, which consisted of 3 phases: habituation, conditioning, and test phases. In the virtual scenario, participants had to explore two rooms in an art gallery, approach virtual humans, and interact with them. To measure the participants' trust toward the virtual humans, their subjective evaluation of the virtual human's trustworthiness and the consequent trust behaviours was recorded. Our study found that both familiar and unfamiliar virtual humans were evaluated as equally trustworthy. Furthermore, participants did not share more information or stand closer, nor did they spend more time in the room with the familiar virtual human. We expected participants to evaluate the familiar virtual human as more trustworthy since they had more information to rely on when making a trust evaluation (Alarcon et al., 2018). An experimental study interested in how familiarity influences trust evaluations found that unfamiliar persons are evaluated as more untrustworthy (Alarcon et al., 2018). It could be that the repeated interaction with the virtual humans in our study fostered a sense of familiarity with both virtual humans (Cochard et al., 2004). It could also be that both virtual humans were evaluated as equally trustworthy regardless of their familiarity since they both had the same body language, facial expressions, and tone of voice.

Furthermore, it is also important to note that not all familiar entities are trustworthy, and not all unfamiliar entities are untrustworthy (Sandstrom et al., 2022). The lack of behavioural differences in our participants contradicts traditional and virtual reality-based experimental studies that indicate a substantial influence of SIP familiarity on trust. For example, repeated interaction with a virtual agent leads to familiarity, self-disclosure, higher levels of trust, and close proximity with the virtual agent (Delton et al., 2011; Gulati & Sytch, 2008; Maloney et al., 2020; Rosenberger et al., 2020; Uzzi & Gillespie, 2002; Vugt et al., 2010). Due to the presence of two virtual humans, it could be that participants made decisions based on their preference for one particular virtual human (based on the virtual human's facial appearance) instead of their behaviour is motivated by trust towards the virtual human. Furthermore, repeated interaction with both virtual humans could have led to a sense of familiarity with both, resulting in no differences in trust.

This is in line with the results of Kiser et al. (2022), which indicated that no significant main effects were observed and no social place preference effect, but only behavioural tendencies concerning the differences in time spent with agents who exhibited happy or angry appearances. This questions the validity of the SCPP paradigm for humans and indicates that it may be unsuitable for measuring trust toward virtual humans in virtual reality in its current state. A validation of the proxies of trust used in this experiment is needed to ensure their sensitivity to differences in trust and their suitability to measure trust toward virtual humans. No prior studies investigated the influence of familiarity on trust with the social conditioned place preference paradigm in virtual reality. This was deemed a meaningful and explorative attempt in this field.

3.4.1. Contribution, Limitations, and Future work

This study has limitations, and the small sample size is one of the various factors that can explain the non-significant results. For example, it could be that the manipulation of familiarity was not strong enough to observe the effect in the subjective ratings and the participants' behaviour, similar to Kiser et al. (2022). Another limitation of the study is that the virtual human's appearance was not evaluated. To control the differences in the virtual human's humanness and trust evaluation the virtual humans were counterbalanced between the two rooms. A survey found that the degree of humanness in information technology (such as agents and social media platforms) positively impacts trust evaluation (Tripp et al., 2011). Qualities such as trustworthiness, attractiveness, and humanness of the virtual humans were not assessed before their inclusion in this study. It would be recommended to conduct a pre-study to assist with the selection of virtual human's appearance to ensure the virtual humans do not differ in appearance.

Participants only received background information and "encountered" the virtual human via a picture and spoken words. Future studies should consider strengthening this manipulation, perhaps by creating a video of the agent or a live Zoom call. An interview with the agent in real life, with guided questions is another option. Kiser et al. (2022) further explain that the study did not have sufficient conditioning trials due to the possibility of simulator sickness, which could also be true for the non-significant results in this study. Using the social conditioned place preference paradigm to study human behaviour in virtual reality is relatively new, thus there is a lack of standardized protocols and procedures for conducting experiments. Furthermore, the SCPP paradigm primarily focuses on short-term behavioural responses. However, understanding the long-term effects of social conditioning in VR is essential for comprehensively studying human behaviour. Investigating the persistence and generalization of conditioned responses over time can provide valuable insights. Conditioning in animal research usually last 3-5 weeks, whereas the conditioning in this paradigm only lasts 15 minutes (Golden et al., 2017; Kummer et al., 2014). The conditioning phase can be extended, but provisions must be made to lower the risk of simulator sickness, which can confound the results. Another limitation of the study is that the paradigm was not validated as a suitable measure of trust toward virtual humans in virtual reality. Furthermore, the proxies of trust were not validated but they were based on previous laboratory findings (Bailenson et al., 2003; Hall & Hall, 1966; Higgins et al., 2022; Kiser et al., 2022; Vugt et al., 2010; Wheeless & Grotz, 1977), thus it is not certain whether trust can be measured through the proposed behavioural proxies. Future studies should consider these limitations and validate the proxies of trust.

This familiarity study has real-life implications, as people in the virtual world can choose to present themselves as "familiar" (e.g., in such a way that others can recognize them: photo-realistic avatar that looks just like them in the real world) or "unfamiliar" (e.g., in an unrecognizable character: using a cartoon-like character that hides their true appearance) to those who know them in real life. Users can choose how they want to visually present themselves in the virtual social environment by choosing from a standard pre-set of self-representations (that do not look like them, e.g., cartoon-like, monsters, robots, animations, or even elves), or users can customize their self-representation, by changing the shape, hair, skin colour and different parts of the body to their liking (Lin & Latoschik, 2022). These customizable avatars are used in social virtual reality applications like RecRoom and AltspaceVR. Users can choose whether they want to present themselves as familiar or unfamiliar to other users while interacting in social virtual reality. This corresponds with the study's primary objective of examining how familiarity influences trust toward virtual humans in virtual reality. Recently the use of (photo)realistic self-representations (a visual representation that looks just like the user in real life) is becoming more and more popular in social virtual reality (O'Brolcháin et al., 2016). As social interaction in the form of avatars is becoming more accepted, it is necessary to understand how users need to appear in virtual reality to be evaluated as trustworthy (for example, familiar or unfamiliar). Imagine meeting a bank consultant in virtual reality to assist you with a transaction. Will you trust a "familiar" bank consultant more or less than an "unfamiliar" one? This example emphasizes the practical implications of this research study for real-world applications. It sheds light on the factors that influence trust in virtual interactions, which can inform the design of virtual environments and guide the development of strategies to enhance trust and facilitate secure transactions in virtual reality settings. Understanding how familiarity with a social interaction partner influences trust can assist with the development of identity management systems in virtual social environments where social interaction through virtual humans is possible, for example, transactions in virtual reality.

3.4.2. Conclusion

In this study, we investigated how the familiarity of a virtual human influences interpersonal trust in a virtual art museum environment. This study used a variant of the social conditioned place preference paradigm that was adapted to humans and implemented in virtual reality. The main outcome measure is the social place preference effect, which is observed by participants' preference for a specific room within the virtual art museum. In the current study, participants interacted with two virtual humans (one familiar and one unfamiliar). We assessed participants' evaluation of the trustworthiness and humanness of virtual humans. In addition, we assessed various behavioural indicators that serve as proxies for measuring trust. These indicators included the proximity of participants to the virtual humans, the extent of information they shared with them, and their response time. By examining these behavioural measures, we aimed to gain insights into the participants' levels of trust and engagement with the virtual humans in the study. The non-significant results from our study raise uncertainties regarding the suitability of the social conditioned place preference paradigm as a behavioural measure of trust. It is possible that the manipulation of familiarity may not have been sufficiently strong to impact participant's evaluations of the virtual human's trustworthiness, resulting in a lack of discernible behavioural differences. Given these findings, there is a need for further assessment and standardization of the social conditioned place preference paradigm and the proposed proxies of trust as valid measurements of trust. In future experimental studies, it is recommended to incorporate a pre-study phase where virtual humans are carefully selected to exhibit medium to high levels of humanness and trustworthiness. This approach will allow for more robust evaluations of trust.

4. Study 2: The Virtual Maze

Validation of the Virtual Maze as a Behavioural Paradigm to Measure Trust in Virtual Reality

Parts of the following section have already been published as:

J. Cronjé, J. Lin, I. Käthner, P. Pauli and M. E. Latoschik, "Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2401-2411, May 2023, doi: 10.1109/TVCG.2023.3247095.

4.1. Introduction

Chapter 2 introduced, conceptualized (2.1.1), and outlined the measurement (section 2.1.3.) of interpersonal trust. For the work of the present thesis, we adopted the definition of interpersonal trust as: “the attitude that an agent will help achieve an individual’s goal in a situation characterized by uncertainty and vulnerability” (Lee & See, 2004, p. 51). It refers to the dyadic relationship between one person and another specific SIP.

The measurement of interpersonal trust between virtual SIPs (virtual humans) requires researchers’ attention. First, the attribution of human characteristics and human behaviour to VHs could impact trust. Such human factors include cooperativeness (Moradinezhad & Solovey, 2021), physical appearance (facial expressions, avatar behaviour, tone of voice) (Machneva et al., 2022), anthropomorphism, and trust resilience (de Visser et al., 2016). 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 interactions with human-to-virtual human interactions. This could result in a better understanding of social responses to VH interfaces and improve interface design (Zanbaka et al., 2007). For example, it has been found that using self-avatars in shared virtual environments can increase trust formation in collaborations, compared to using only the model of VR controllers as representation (Pan & Steed, 2017). Lastly, compared to people in real life, VHs 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 (Lin & Latoschik, 2022), 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 toward VHs a priority.

Several measures to assess generalized trust (how one trusts others in general) have been previously proposed (see Chapter 2, section 2.1.3.). However, there is a lack of instruments that measure specific trust (trust towards a specific person in a specific circumstance). This holds particularly true for paradigms that measure interpersonal trust towards VHs in VR. Hale et al. (2018) proposed a virtual maze task that relies on advice-seeking behaviour to measure trust towards two opposing characters (see Chapter 2, section 2.3. for a complete description of the paradigm). However, in its proposed form, the maze task is unsuitable for measuring interpersonal trust towards VHs as SIPs 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.

The presented work addressed these points. The motivation for an improved version has the VR community in mind, and it is rooted in the need for an in-the-moment interpersonal trust measurement tool that considers fundamental inherent VR characteristics. This tool should be sensitive to the manipulation of the VH's trustworthiness that ultimately guides trusting behaviour. Our contribution includes a paradigm for investigating interpersonal trust towards VHs (agents and avatars) in VR. Compared to the paradigm of Hale et al. (2018), our VHs have more realistic social cues, and our task emphasizes the immersion and interactivity of VR. Our paradigm includes one specific VH during the maze task. Participant's behavioural differences stem from their subjective evaluation of the VH instead of comparing two VHs, 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 amid uncertainty.

Building on the work of Hale et al. (2018) (see section 2.3.2.) and considering the limitations of their paradigm, 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. We conducted a validation study to verify whether our paradigm can measure differences in interpersonal trust toward VHs.

Improvements to the Virtual Maze

The virtual maze of Hale et al. (2018) 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 toward VHs. However, certain adaptations are needed. First, VHs include virtual agents and avatars, which are usually considered the proxies of AI or humans. Considering the definition of interpersonal trust, VHs, 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. (2018), participants perceived the trustees as pre-scripted characters that reflect the preliminary design of the experiment and the will of the experimenters. This could 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. (2018). 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 VR, 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, a head-mounted display (HMD), and a desktop PC, respectively (Hale et al., 2018). In either version, participants can only navigate with a joystick to approach the virtual characters. Compared with 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 (Jung et al., 2017; Roth & Latoschik, 2020). Given the often-reported effects of virtual embodiments and presence on secondary factors such as emotional response (Tsankova et al., 2013) and especially trust (Salanitri et al., 2016), 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. (2018): 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. 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 SIPs rather than being limited to pre-scripted agents and behaviour is 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.

4.1.1. Hypotheses

The reviewed literature inspired the following hypotheses:

H1: The manipulation of trustworthiness is successful; participants 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 ask for advice significantly more often than those in the untrustworthy condition.

H3: Participants in the trustworthy condition follow advice significantly more often than those in the untrustworthy condition.

H4: Participants in the trustworthy condition respond quicker to advice than those in the untrustworthy condition.

For the *trust game,* we expected the following:

H5: Participants in the trustworthy condition invest more tokens than in the untrustworthy condition.

4.2. Methods and Materials

4.2.1. Study Design and Participants

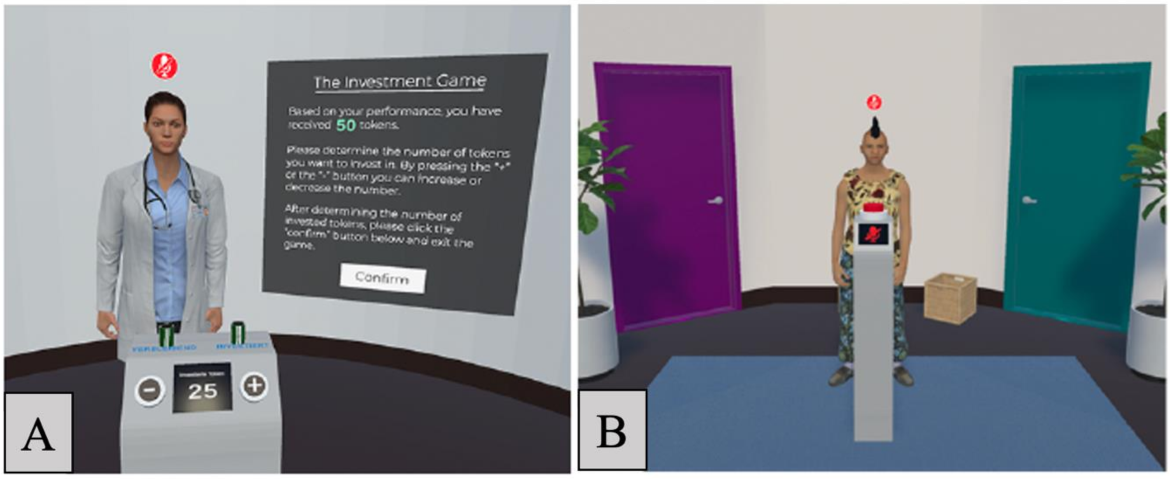
We expected a large effect size for the outcome measures based on the work of Hale et al. (2018). We recruited 70 women (*M*: 23.12 years, *SD*: 4.31, range 18 and 35) without psychiatric or neurological disorders. Participants registered to participate in the experiment via the university’s SONA system. In this between-subjects design, half of the participants participated in the trustworthy condition, and the other half participated in the untrustworthy condition. Every participant was compensated with 15 euros for participating in the experiment, which lasted approximately 1.5 hours. 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 (Eagly & Wood, 2012; Rau, 2012; Wu et al., 2020).

4.2.2. Apparatus

The application for the experiment was implemented as an immersive VR experience using Engine, 2020 Unity.3.14f 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 Head-Mounted-Display (HMD) and two controllers connected to the PC through the Oculus Link service.

Figure 10: Virtual Environments: Virtual Maze and Trust Game



Note: A) the trust game B) the virtual maze

Virtual Environment. The virtual environments were created with Blender 2.92 (Blender Foundation), and assets were downloaded from the 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 trust game can be seen in Figure 10.

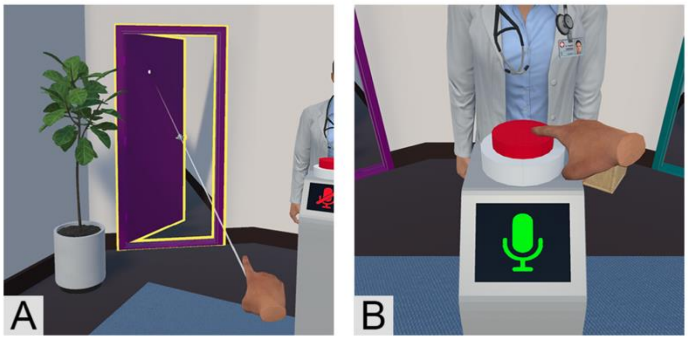
The Introductory Video was an essential component of our paradigm. It served three purposes: 1) to foster the idea that a real person was behind the VH to ensure participants perceived the trustee as an interaction partner; 2) to inform participants that there were multiple escape routes and that the trustee could provide advice according to where they were located in the maze at that moment; 3) to emphasize that the trustee had a choice to either be helpful or misleading to prevent participants from blindly following all the given advice.

The Trust Game is an established, validated behavioural measure of trust (Berg et al., 1995). We used a variant of the trust game in which, in the beginning, participants received 50 tokens. They could decide how many tokens out of 50 they wanted to invest by sending them to the trustee (see Figure 10A). 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 was that the more tokens participants invested, the higher the trust in the trustee. The trust game also served as an additional measure for the behavioural proxies of trust in the maze task.

The Virtual Maze consisted of several rooms, and in each room, there were two doors opposite the participants, which led to different rooms. As illustrated in Figure 10B, the trustee avatar appeared in the middle of the room, as it does not imply preference of which door to choose. Participants could walk around freely in the room. To open a door, participants had to point to the door with a controller and press a button on the controller (Figure 11A). When they decided to ask for advice, they needed to approach the avatar and press the red button in front of it with their virtual hands to unmute them (Figure 11B and Figure 12). It was simpler and quicker in operation to directly open a door than to ask the avatar for advice by design. Thus, it was less likely that participants blindly sought advice all the time. When the trustee was unmuted, the status symbol turned green, and the trustee could talk to the participants to give advice. The pre-recorded audio with advice was made to sound natural and realistic and included hesitant pauses and filler words such as "uhm...". Participants were informed that the trustee could not hear them to avoid attempts at conversations.

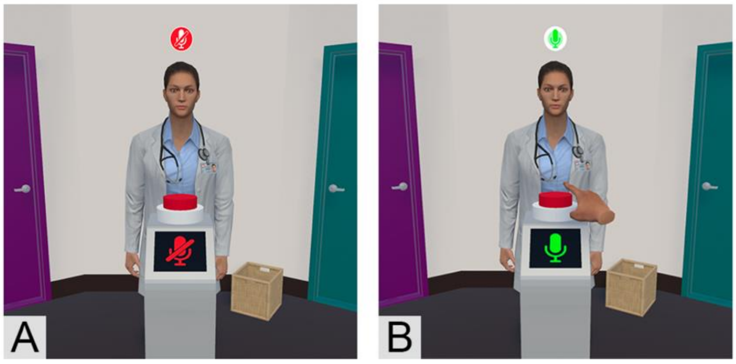
Once participants selected a door, they would be teleported to a new room. To give participants the impression that they were entering a new room at each trial, minor changes to the virtual environment were made (for example, the colour of the doors changed, plants or objects were added and/or removed, etc.). The trustee avatar would appear muted again in the middle of the room. Unknown to participants, there was no right or wrong door to choose from. After participants decided to enter a new room 12 times (12 trials), they were informed that they had escaped the maze. Since participants received no feedback on their progress in the maze, their experience of uncertainty was increased, and their decision-making was uninfluenced by external motivators.

Figure 11: Virtual Environment: Virtual Maze



Note: A) Participants could point to a door and open it by pressing a button on their controllers. B) To seek advice from the virtual human, participants approached and unmuted the avatar.

Figure 12: Unmute the Virtual Human.



Note: A) The virtual human was muted whenever participants entered a new room. B) Participants could press the red button in front of the virtual human to unmute them if they wanted to ask for advice.

Virtual Human. An online pre-study was conducted to assist in selecting the appearance of the two VHs: one for the trustworthy condition and one for the untrustworthy condition. Six avatars were selected from the Rocketbox library (Gonzalez-Franco et al., 2020), and six avatars were created with MakeHuman 1.2.0. 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 (M = 6.08, SD = 0.97) and the most untrustworthy female avatar score (M = 2.30, SD = 1.02) were selected for the main experiment (see Figure 13).

Figure 13: Virtual Humans

high- trustworthiness *low- trustworthiness*

Note: A) most trustworthy female avatar, B) most untrustworthy female avatar.

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 SIPs. Thus, we ensure the realism of the VH in several aspects. The avatar stood in an idle pose, with its head and eyes following the participants naturally. While speaking (a pre-recorded audio was played), the mouth of the avatar moved accordingly through the Oculus Lipsync plugin (*Oculus Lipsync for Unity Development: Unity*). Our application supports avatars compatible with humanoid unity skeletons with facial blend shapes.

4.2.3. Trustworthiness Manipulation and Measurements

Manipulation of Trustworthiness

The trustworthiness of the VHs was manipulated, and untrustworthy 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. The manipulations included the avatars' appearance, tone of voice, and engagement during the task (Table 6). We used pre-scripted agents to ensure that all participants were exposed to the same contents and interactions instead of having real people control the avatars. However, participants were under the impression that they were interacting with real humans.

Table 6: Manipulation of the Virtual Human's Trustworthiness

	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

Measurement of Trust

To assess the success of the manipulation, participants reported their subjective evaluation of the VH’s trustworthiness on 7 point-Likert scale ranging from 0 - *not trustworthy at all* to 7 - *completely trustworthy*. Furthermore, we assessed whether the following behaviours are sensitive to the differences in interpersonal trust (these were also the main outcome measures).

Main Outcome Measures

1) How often participants ask for advice, 2) how often participants follow advice, and 3) the response time of following the advice. As an additional behavioural measurement of trust, two trials of the one-shot trust game were played (before and after the maze task). We measured the 4) number of tokens participants invested in each trial.

Questionnaires

The *pre-VR measures* participants reported demographic information and health questions. The *post-VR measures* included explicit questions about their evaluation of the avatar (humanness, friendliness, realness, helpfulness, trustworthiness) during the VR tasks. Furthermore, the IPQ (Schubert, 2003) (see Chapter 3 for a complete description of the IPQ), The Social Presence (SP) Questionnaire (Bailenson et al., 2001), was used to examine to what extent participants perceived the avatar as a real human being and SIP, participants had to indicate the extent to which they felt the statements regarding the presence of the SIP are true, on a 7-point scale. With – 3 indicating “do not agree at all”, 0 “neither agree nor disagree”, and 3 “totally agree”. Furthermore, the SSQ (Kennedy et al., 1993) (see Chapter 3 for a complete description). Lastly, the KUSIV3 (Beierlein, Kemper, Kovaleva, et al., 2012) was included. In 2012, the Interpersonal Trust Scale of Rotter was adapted and reduced to a 3-item scale, developed in German. This shorted scale, known as the *"KUSIV3"* (English and German), is highly economical for capturing the psychological trait of interpersonal trust. In this 3-item scale, the KUSIV3 measures general trust in strangers or close acquaintances (Beierlein et al., 2012). Participants indicate general trust on a 5-point rating scale ranging from 1 "strongly disagree" to 5 "strongly agree". The subjective measurements were implemented with LimeSurvey 4.5.0 (LimeSurvey Development Team, 2012). Table 7 summarize the mean and standard deviation of the subjective measurements used in this study for the trustworthy and untrustworthy condition.

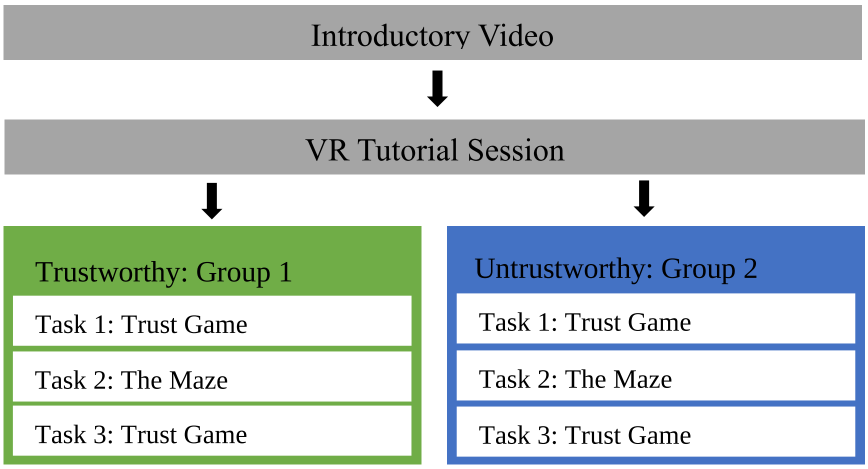
Table 7: Descriptive Questionnaire Data

	Group 1: Trustworthy VH		**Group 2: Untrustworthy VH**	
	M	**SD**	**M**	**SD**
	Post – VR Measures			
Age	22.91	2.92	23.32	3.34
IPQ Spatial Presence	3.22	1.06	3.31	1.01
IPQ Involvement	2.17	0.81	2.28	0.83
IPQ Experienced Realism	1.41	0.56	1.45	0.57
IPQ General	4.74	1.66	4.85	1.76
SP - in presence of another	0.36	1.92	0.43	1.90
SP- feel watched	-0.04	2.01	0.32	1.99
SP - other is not real	0.40	2.14	0.30	2.05
SP - other seem alive	0.11	1.74	0.06	1.79
SP - other is technical device	-0.15	1.84	-0.02	1.89
SSQ Total	0.23	0.26	0.24	0.21
SSQ Nausea	0.12	0.23	0.14	0.19
SSQ Oculomotor	0.30	0.37	0.32	0.29
SSQ Disorientation	0.26	0.33	0.25	0.31
KUSIV3	3.41	0.47	3.48	0.42

Note: VH = Virtual Human; IPQ = Igroup Presence Questionnaire; SP = Social Presence Questionnaire; SSQ = Simulator Sickness Questionnaire; KUSIV3 = Shortened German Version of the Interpersonal Trust Scale of Rotter

4.2.4. Procedure

Figure 14: Experimental Procedure



Note: VR= Virtual Reality, VH = Virtual Human. A summary of the experimental procedure of the validation study. Half of the participants in group 1 interacted with the trustworthy VH, and the other half (group 2) interacted with the untrustworthy VH.

At the beginning of the session, participants read the study information (see Appendix D), provided informed consent (see Appendix B), and filled in the *pre-VR measures* (demographics and VR-related health questions). The *introductory video* was played after the pre-VR questionnaires were completed. After that, participants were guided by the experimenter to put on the HMD and enter the *VR interaction tutorial*. Thereafter, participants entered the *testing room*, where a connection with the trustee was established, and they could see the avatar for the first time. In the testing room, participants were asked to unmute the trustee to enable the trustee to talk to them. The testing room was used to ensure that the connection had been established between the participant and the alleged avatar. Participants then started the first experimental task, the VR *trust game* (*first trial)* (see Figure 10A). Then they were teleported to the second VR task, which consisted of our version of the *virtual maze task* (see Figure 10 B). 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 received the same advice (if they asked for it) within each condition, and the avatar displayed the same behaviour. After completing the maze task, the *trust game* with the same setup was played (*second trial*). Participants then removed the HMD to fill in the post-VR questionnaires. Figure 14 briefly summarizes the procedure of the experiment.

4.2.5. Statistical Analysis

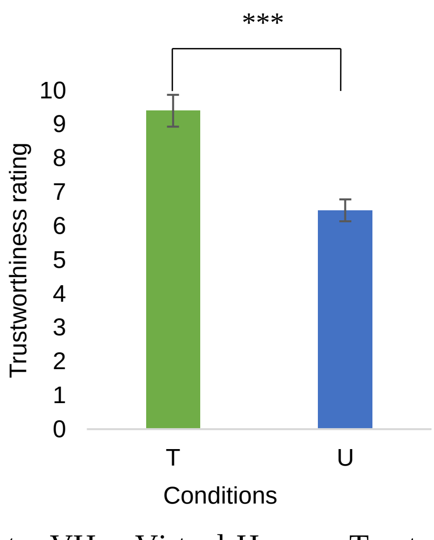
Jamovi version 2.3.21.0 (The Jamovi project, 2021) was used to conduct t-test unless the requirements were not met, in which case non-parametric tests were chosen for hypothesis testing. Independent sample t-tests were conducted to test the scores of these subjective measurements between participants as well as behavioural measurements between the trustworthy and untrustworthy conditions. The level of significance was set at $p < .05$.

4.3. Results

Manipulation Check

Participants in the trustworthy condition evaluated the avatar as significantly more trustworthy ($M = 9.40$, $SD = 2.28$) compared with participants in the untrustworthy condition ($M = 6.46$, $SD = 2.86$), $t\ (68) = 4.76$, $p < 0.001$, $d = 1.14$ (see Figure 15). The results confirmed that our manipulation of trustworthiness was successful, and H1 can be accepted. The two groups did not differ in their propensity to trust others in general, they were equally trusting, $t\ (68) = -1.64$, $p = 0.107$, $d = -\ 0.391$.

Figure 15: Manipulation of VH Trustworthiness



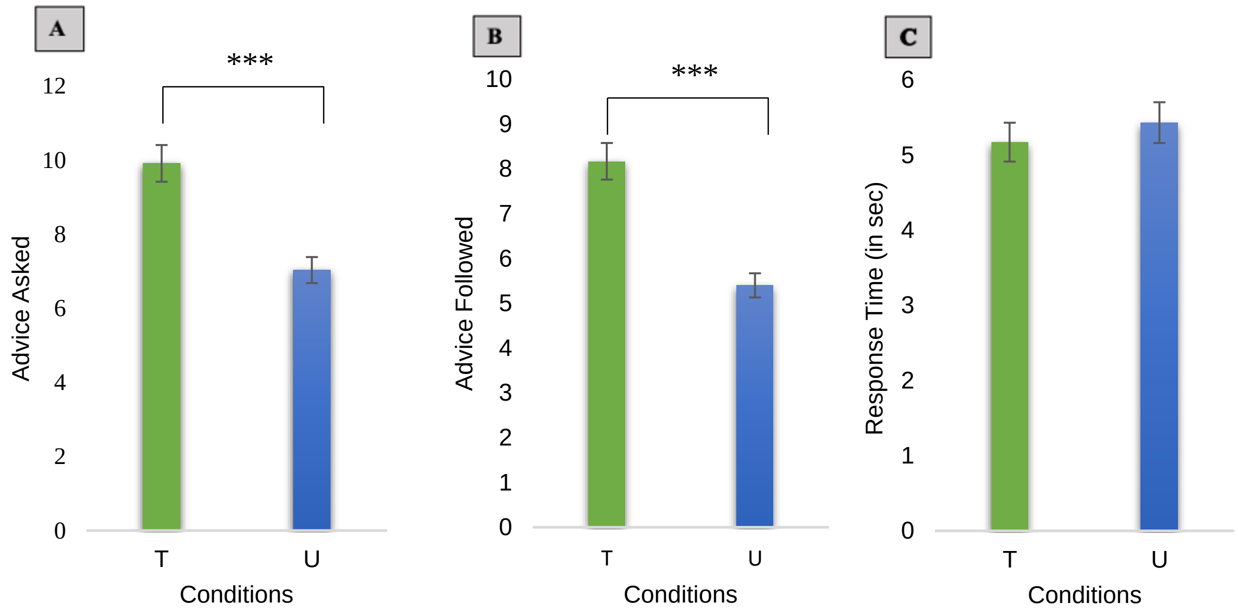
Note: VH = Virtual Human, T = trustworthy, U = untrustworthy, the error bars indicate the standard error, *** $p < .001$.

Subjective Measures

We 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 = 0.47, 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.

Behavioural Measures

Figure 16: Main Outcome Measures Between Conditions



Note: T = trustworthy, U = untrustworthy, the error bars indicate the error bars indicate the standard error, *** $p < .001$.

Advice Seeking

Participants in the trustworthy condition asked for more advice (*M* = 9.91, *SD* = 2.54), compared to participants in the untrustworthy condition (*M* = 7.03, *SD* = 2.96), *t* (68) = 4.48, *p* < 0.001, *d* = 1.05 (Figure 16A). Thus, H2 can be accepted.

Advice Following

Participants in the trustworthy condition followed more advice (*M* = 8.17, *SD* = 2.79) compared to those in the untrustworthy condition (*M* = 5.40, *SD* = 2.80), *t* (68) = 3.93, *p* < 0.001, *d* = 0.94 (see Figure 16B).

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 was observed for participants in the trustworthy condition (*M* = 0.82, *SD* = 0.12) to follow advice more often than in the untrustworthy condition (*M* = 0.78, *SD* = 0.17). 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 that was asked.

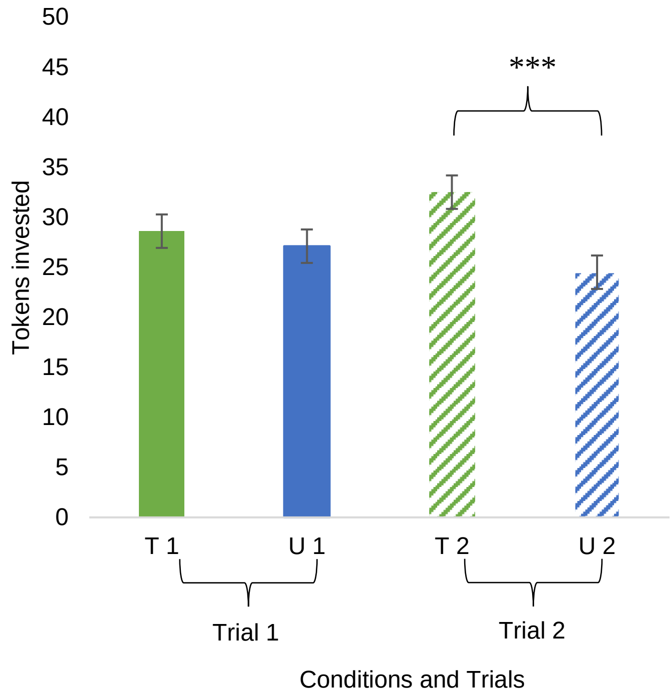
Response Time to Execute Given Advice

The requirements for a t-test were not met, thus a non-parametric U-test was conducted. 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, *U* (68) = 528, *p* = 0.326 (see Figure 16C). Thus, participants in the trustworthy condition did not respond faster (selecting a door), and H4 was rejected.

Trust Game

Investment of Tokens Between Conditions.

Figure 17: Tokens Invested in Trust Games.



Note: T = trustworthy, U = untrustworthy. Participants were given 50 tokens at the start of each trust game; the graphs depict the total number of tokens invested and the error bars indicate the standard error, *** $p < .001$. The solid bars depict represent the first trust game and the striped bars the second trust game.

In the first trial before the maze task, participants in the trustworthy condition (M = 28.6, SD = 7.51) did not invest significantly more tokens than in the untrustworthy condition (M = 27.1, SD = 9.84), U (68) = 548, p = 0.445. However, significant differences were observed, U (68) = 317, $p < 0.001$, 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) (see Figure 17). Thus, H5 can be partially accepted. Only in the second trial of the trust game after the maze task were more tokens invested in the trustworthy condition. The trials had no significant effect on the number of tokens invested, $F(1,68) = 0.34$, $p = 0.563$, $\eta_p^2 = 0.005$. There was a significant difference between the number of tokens invested between the conditions in the second trial (M = 28.5, SD = 10.00) compared with the first trial (M = 27.5, SD = 8.72). There was a significant difference in the number of tokens invested in the trustworthy condition (M = 30.5, SD = 6.98) compared with the untrustworthy condition (M = 25.8, SD = 8.54), $F(1,68) = 6.47$, $p = 0.013$, $\eta_p^2 = 0.087$.

4.4. Discussion

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 trust 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 can be a good proxy for measuring trust. 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 (Peña & Yoo, 2014; Surprenant, 2012) and vocal cues (Torre, 2017; Tsankova et al., 2013) have an impact on trust; our results indicated that the manipulation of these factors does indeed influence the evaluated trustworthiness of virtual humans.

In the trustworthy condition, the advice that was given was followed significantly more compared with the advice followed in the untrustworthy condition. These results of our study should not be compared to the study of Hale et al. (2018), which detected significant differences in the ratio of advice following. In their design, participants could always ask for advice from two characters that give opposing suggestions, and they could only choose one to follow; in our between-subject design, participants had 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 (Yamagishi & Yamagishi, 1994) score.
Despite the descriptive differences, no statistical difference in response times (to follow advice) was observed in both conditions. The tendency of shorter response times in the trustworthy condition could be due to participants' trust towards the avatar and a feeling of comfort with them. This results in participants' quick responses, without spending a lot of time thinking through their options. According to a content analysis that investigates factors that influence perceived trustworthiness and risk-taking behaviours, prompt and predictable responses are highly correlated with trust (Breuer et al., 2020).

In the first trial of the trust game (Berg et al., 1995), no difference was detected in the amount of tokens invested 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 trust 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 trust game to be a tool for assessing trust building during social interactions with virtual humans.

4.4.1. Contribution, 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 (Ben-Ner & Halldorsson, 2010) and tend to predict external behaviour poorly (Armitage & Christian, 2003; McCambridge et al., 2012). 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 trust 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 the generalized trust.

In previous studies that measured trust towards virtual agents or avatars (Bente et al., 2008; Hale et al., 2018; Pan & Steed, 2016), 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 (Moradinezhad & Solovey, 2021) instead of interpersonal trust. In our paradigm, we claim that our framing was successful as participants were under the impression that they were interacting with a real person. Several participants explicitly mentioned that they really thought there was another participant after being debriefed that they were, in fact, interacting with pre-scripted agents. Therefore, they were more likely to consider virtual humans as social interaction partners, which is a prerequisite for using this paradigm in studying avatars in social virtual reality. 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 virtual reality 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 virtual reality. Thus, our paradigm shows greater potential for relevant studies in virtual reality, where fundamental virtual reality characteristics (such as embodiment (Roth & Latoschik, 2020), virtual body ownership (Slater, 2008), presence (Schubert, 2003), congruency and plausibility (Latoschik & Wienrich, 2022)) 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 virtual reality in a natural scenario (e.g., to measure whether a social virtual reality user perceived a passing stranger as an avatar trustworthy or not). Instead, our variant of the virtual maze provides the academic community with a flexible foundation to build on for future investigations of trust. In the original paradigm, it is impossible to manipulate external factors that influence trust, such as cognitive load (Ahmad et al., 2019; Samson & Kostyszyn, 2015), as two virtual characters should always be included. Our modifications also provide flexibility for between-subject designs, often preferred to avoid social desirability bias (Agarwal et al., 2011; Alsulaiman & Saddik, 2006; King & Bruner, 2000; O'Brolcháin et al., 2016; Olade et al., 2020; Sluganovic et al., 2016). 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 supposed map of the maze).

Despite the advantages mentioned above, we have identified some limitations and directions for further research. The impact of the colours of the doors in each room was not assessed, it could be that participants' preference for a certain colour could have influenced their decision on which door to choose. Previous experimental research showed that colour influences decision-making (Silic & Cyr, 2016). Furthermore, 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 virtual reality applications, users' avatars can already reproduce body movements and even facial expressions with additional trackers (Lang, 2022; Melnick, 2022). 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 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 study 2.

4.4.2. 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 by these tools. To date, no subjective or behavioural measurement tool (apart from Hale et al.'s novel maze paradigm (Hale et al., 2018)) 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 virtual reality and considers fundamental virtual reality 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 (study 2) 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.

5. Study 3: What you do

The Impact of Cognitive Load on Trust Towards a Virtual Human in Virtual Reality

5.1. Introduction

Trust is essential, especially when high degrees of uncertainty and cognitive load (CL) are experienced. In the Cognitive Load Theory, Sweller defined CL as the degree of demand placed on the working memory during a task (Sweller, 2011). The theory explains that working memory possesses limited capacity, and when CL surpasses this capacity, it can lead to cognitive overload, hampering effective information processing (Sweller, 2011). Furthermore, CL has a negative impact on cognition and it decreases performance in all types of tasks (Adcock, 2000; Hinson et al., 2002; Rydval, 2012). For example, decision-making strategies are negatively affected under high CL, as people are more impulsive and less analytical (Duffy & Smith, 2012; Hinson et al., 2002), they tend to neglect information (Gilbert et al., 1988; Swann et al., 1990), are prone to use decision heuristics (Bohner et al., 2002), and resort to cognitive biases (Gilbert, 1989; Hernandez & Preston, 2013). CL also decreases self-control (Shiv & Fedorikhin, 1999; Ward & Mann, 2000) and the willingness to take risks (Benjamin et al., 2013). Under conditions of high CL, fewer cognitive resources are available to critically think about a decision, this makes analytical processing more challenging and leads to cognitive shortcuts.

These consequences of CL are noticeable in the advancement of communication technologies on users' cognitive capacity and functioning (Samson & Kostyszyn, 2015). In modern societies, most interactions and transactions take place online and include: human-to-human and human-to-computer/machine interactions (Lieberman & Schroeder, 2020). Users engage with the system or other users (online or offline), while simultaneously maintaining a level of situational awareness (Lopes et al., 2018). The increase in multitasking and the availability of endless information leads to a state of constant cognitive overload (Klingberg, 2009). The more complex a task, or the more complex social interactions become, the more critical role trust plays according to traditional experimental studies (Bagneux et al., 2012; Myers & Tingley, 2016). Samson & Kostyszyn, (2015), suggest that trust is a process that has both social and cognitive components. The trust process is reliant on confidence in another's intentions, as well as a rational calculation of a specific person, in a specific context (Becker, 1996; Benjamin et al., 2013; Rempel et al., 1985). An experiment study investigated the impact of CL on trust behaviour between two persons, using the trust game (Samson & Kostyszyn, 2015). The results showed that CL is a key factor that decreases trust toward an SIP and increases impulsive decision-making. The success of all our relationships is reliant on interpersonal trust (online and offline).

Social VR is an example of communication technology in modern society where users' cognitive resources and trust are challenged. In Social VR, users can interact with other users and the environment (Nordin Forsberg & Kirchner, 2021). Users enter the virtual world in the form of a virtual character (with the appearance tailored to their liking) and interact with other virtual characters that are either human-controlled (agents) or computer-controlled (avatars). The social interaction between VHs has similar social norms as human-to-human relationships (Nass et al., 1994). Therefore, interpersonal trust towards VHs in VR and the psychological factors that influence trust evaluation need researchers' and VR designers' attention. There is limited research on how CL influences trust evaluations of VHs in VR therefore, we try to understand this relationship from the literature on human-automation/machine use. We are interested in investigating how participants' experience of CL influences trust towards VHs in VR. Various factors influence our evaluation of trust in people, systems, or technology. For example, the demand for cognitive resources, how the interaction partner sounds or appears, and their intonation (Gupta et al., 2019). Furthermore, CL is crucial when designing and developing educational games (Brunken et al., 2003; Plass et al., 2010). Analytical processes are usually more difficult to perform under high cognitive load (HCL). HCL thus leads to decreased analytical thinking and decision-making strategies (Ahmad et al., 2019; McBride et al., 2011; Samson & Kostyszyn, 2015), resulting in more reliance on external resources even if trust is low (Chen et al., 2016).

Several studies investigated the effects of CL on trust. An experimental study investigated the impact of CL on trust towards a robot in a collaborative task (Ahmad et al., 2019). They found that the experience of high degrees CL during emergencies or critical situations might have a negative impact on trust towards the human or computer (i.e., SIP) (Ahmad et al., 2019). Furthermore, they found that an increase in CL decreases a participant's subjective trust evaluation and trust behaviour towards the SIP (Ahmad et al., 2019). If the participants trust the SIP, then they may stop to critically think through their options and the consequences of their decisions, which results in a decrease in the experience of cognitive load (Ahmad et al., 2019). Samson & Kostyszyn, (2015) found that participants express significantly less trust in a Trust Game and act more impulsive during HCL. Zhou et al. (2017) investigated the effects of uncertainty and CL on users' trust in an artificial intelligent (AI) agent in predictive decision-making. They found that high uncertainty led to increased trust, only under low cognitive load (LCL) conditions when users had adequate cognitive resources to process the information.

When cognitive resources were low (under conditions with HCL), uncertainty decreased trust in the AI and its suggestions. Peña & Yoo (2014) considered the effects of avatar appearance (clothing) and CL on virtual social interactions in a virtual store, such as participant attitudes, trust, bidding, and interpersonal distance. They found that avatar salespeople dressed in white clothing are perceived as more trustworthy and persuasive, leading to participants standing closer to them. Furthermore, when participants were under HCL, both avatars (in white and black clothes) were trusted equally (Peña & Yoo, 2014).

Previous experiments in VR that measured trust towards VHs (avatars or agents) used a combination of subjective (primarily self-report questionnaires) and objective measures (behavioural measures and proxies of trust) (Aseeri & Interrante, 2021; Hale et al., 2018; Pan & Steed, 2016). In study 2 (Chapter 4 or Lin et al. 2023) we modified and validated a variant of the virtual maze paradigm from Hale et al., (2018) that measures interpersonal trust towards VHs as SIPs in VR. In this paradigm, participants allegedly need to escape from a virtual maze by entering as few rooms as possible. In each room, a VH is available to provide guidance on which door to choose. Participants can decide whether to ask for advice or not and whether or not to follow it. These behaviours serve as proxies of trust.

The present study employed the maze paradigm to investigate the impact of CL on trust towards a VH. CL was manipulated with an auditory version of the n-back task (del Angel et al., 2015), as a secondary task. Based on previous findings outside of VR, we expect that HCL leads to lower trust ratings and fewer trust behaviours (Ahmad et al., 2019; Chen et al., 2016; McBride et al., 2011; Samson & Kostyszyn, 2015).

5.1.1. Hypotheses

The reviewed literature inspired the following hypotheses:

H1: Participants evaluate the SIP as less trustworthy in the HCL condition.

H2: Participants have fewer advice-seeking behaviour in the HCL condition.

H3: Participants have fewer advice-following behaviour in the HCL condition.

H4: Participants respond to advice faster in the HCL condition.

For the *trust game*, we expected the following:

H5: Participants who evaluate the avatar as trustworthy invest more tokens.

5.2. Methods and Materials

5.2.1. Study Design and Participants

We aimed to reveal at least a medium effect size (d = 0.5, β = 0.90, α = 0.05). A calculation in G*power (Faul et al., 2007) indicated that 44 participants would be sufficient. We decided to recruit 50 women (*M* = 23.28 years, *SD* = 3.34, range: 18 and 35) without psychiatric or neurological disorders. Participants registered to participate in the experiment via the university's online recruitment system (SONA), and they received 10 euros as compensation (the experiment lasted 1 hour). The Ethical Committee of the Psychological Institute of the Faculty of Human Sciences of the University of Würzburg reviewed the protocol and provided ethical clearance (GZEK 2021-70). In a within-subjects design, each participant took part in an HCL and an LCL condition; the order of conditions was counterbalanced across participants.

5.2.2. Apparatus

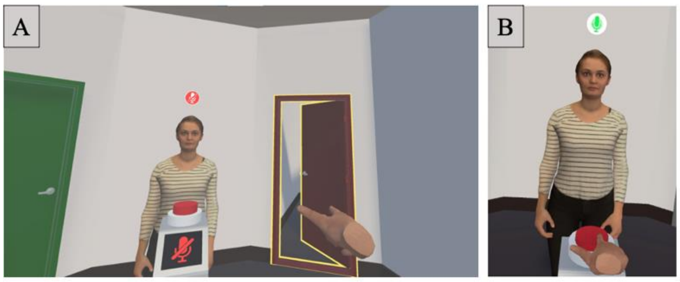
The experiment application was implemented as an immersive VR experience using Unity Engine 2020.3.14f and ran on a VR-capable PC (Intel Core i7-2600K, Nvidia GTX 970 16GB RAM). The VR hardware consisted of an Oculus Quest 2 HMD and two controllers connected to the PC through the Oculus Link service.

Virtual Maze paradigm (see Chapter 4, study 2 or Lin et al. 2023 for a complete description). Participants were told to navigate through a virtual maze by entering as few doors as possible. The maze task included several measures to strengthen the belief that participants interacted with an avatar controlled by another participant; for details see Chapter 4 or Lin et al., 2023. On their way through the maze, participants were able to interact with a trustee (VH) and the virtual environment. In each room, participants faced two closed doors opposite them and needed to choose which door to enter next (see Figure 18). Participants could choose to ask the VH for advice, by pressing a button to unmute the VH. If advice was asked, the pre-scripted agent advised the participant with a pre-recorded audio to choose a particular door. Upon choosing a door by a pointing gesture and confirmatory button click, participants were teleported to a new room. After they entered 12 rooms, participants were finally told that they had escaped the maze. To ensure comparability between conditions, participants heard the same advice from the same VH (supposedly an avatar) but in a different order between conditions.

Main Outcome Measures

The following measures served as proxies of trust and were recorded as the main outcome measures of the current study during the maze paradigm: 1) the number of times participants request advice, 2) the number of times advice is followed, and 3) the time it took to execute the advice (time from button press to ask advice to the selection of the advised door).

Figure 18: The Virtual Maze Task.



Note: A) Participants must select one of the two doors in front of them to navigate through the maze. B) To ask the virtual human for advice (allegedly controlled by a real person), participants must press the red button to unmute the virtual character. To avoid conversational attempts, participants were informed that the trustee could not hear them but only gave advice when unmuted.

The Trust Game (TG) (by Berg et al., 1995, as described in Chapter 4). We used a one-shot variant of the Investment Game (Lin et al., 2023), in which participants received 50 tokens and could decide how many tokens out of 50 they wanted to invest by sending them to the trustee (see Figure 19). The participants were told that the invested amount would be multiplied by four, and the trustee could then return half of the tokens or none (participants were told they would receive feedback at the end of the experiment). The number of tokens invested served as a further outcome measure.

Figure 19: The Trust Game

Note: Participants can invest up to 50 tokens. The number of tokens invested in the virtual human (supposedly an avatar) served as a proxy of trust.

Virtual Human. The female avatar was selected in a pre-study (N = 50) and was rated to have a medium attractiveness and high realism score (Lin et al., 2023). The current study also used an agent to ensure that only CL has changed across the conditions, as in study 2 (Chapter 4 and Lin et al. 2023). At the start of the virtual maze, the VH stood in an idle pose, with its head and eyes following the participants in a natural way, and its mouth moved according to its speech. The VH used in this experiment was a 3D scan using the fast generation of a realistic avatar (Achenbach et al., 2017). Participants completed the maze task with the same VH in LCL and HCL conditions. To ensure comparability and consistency between conditions, participants heard the same advice from the same VH (supposedly an avatar) but in a different order (for example, recording 1-12 in the one condition and then 12-1 in the other condition).

5.2.3. Cognitive Load Manipulation and Measurements

Manipulation of Cognitive Load

To increase CL in the HCL condition, participants had to perform a secondary task in addition to the maze task - an auditory version of the one-back task (del Angel et al., 2015). Beeps of different frequencies (250 Hz, 450 Hz, or 650 Hz) were presented via the HMD over the right ear (for 1.5 seconds with 2 seconds of silence between beeps). Advice from the VH was played over the left speaker of the HMD. During the HCL condition, participants were asked to press a button (on the left controller) when two consecutive beeps were identical. The same beeps were played in the LCL condition, but participants were told to ignore them.

To control order effects, half of the participants started with either the HCL condition (escaping from the maze while doing the secondary auditory version of the 1-back task simultaneously), and the other half started with the LCL condition (escaping from the maze with no additional task).

Measurement of Cognitive Load

To assess the success of the manipulation, participants rated the mental effort during the maze task after each condition on a 10-point scale ranging from 1- *not demanding at all* to 10 – *very demanding*.

Furthermore, the participant's behavioural performance on the n-back task (in the HCL condition) was assessed. A button press up to 1.5 seconds from the beginning of the beep is considered a "hit"; anything after 1.5 seconds is regarded as a "miss." Participants listened to navigation advice from the VH through the left side of the HMD and simultaneously the one-back task auditory task in the right ear of the HMD while navigating through the maze during the HCL condition.

Questionnaires

Table 8: Descriptive Questionnaire Data

	Pre – VR Measures	
	M	**SD**
Age	23.16	3.24
BFI Extraversion	3.27	1.00
BFI Agreeableness	3.47	0.89
BFI Conscientiousness	3.45	0.54
BFI Neuroticism	3.06	0.53
BFI Openness	3.69	0.84
	Post – VR Measures	
GTS	4	0.66
KUSIV3	3	0.51
STAI State	1.45	0.35
SSQ Total	11.74	13.50
SSQ Nausea	7.63	11.24
SSQ Oculomotor	14.71	15.94
SSQ Disorientation	22.27	32.94
IPQ Spatial Presence	3.91	0.79
IPQ Involvement	3.35	0.77
IPQ Experienced Realism	2.75	0.61
IPQ General	3.92	1.19

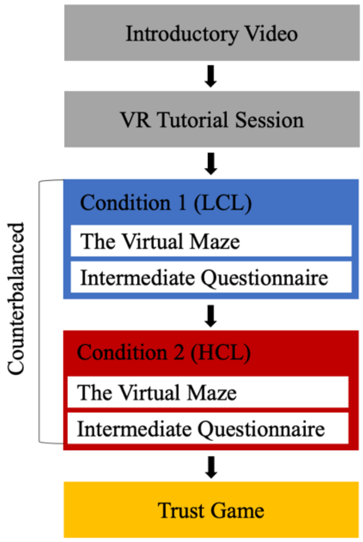
Note: CL = Cognitive Load; BFI = Big Five Inventory; GTS = General Trust Scale; KUSIV3 = Shortened German Version of the Interpersonal Trust Scale of Rotter; STAI State= State Anxiety Inventory; SSQ = Simulator Sickness Questionnaire; IPQ = Igroup Presence Questionnaire.

The pre-VR measures participants reported demographic information, general health questions, BFI personality assessment (Rammstedt & John, 2007), and the State Anxiety questionnaire of the STAI (Laux, 1981) (see Chapter 3 for a full description of the BFI and STAI). *The intermediate measures* between conditions include a self-constructed questionnaire to assess participants' experience of mental effort (manipulation check), the enjoyability of the task, explicit trust towards the avatar, perceived helpfulness of the avatar, and their trust towards the person controlling the avatar in real life on a 10-point scale (ranging from 1- *not at all* and 10 - *completely*). For example, "How much mental effort was required in taking in and processing information?" and "Do you trust the virtual human?" (A rating higher than 5 out of 10 indicated that the participant trusted the avatar). *The post-VR measures* included the

SSQ (Kennedy et al., 1993) and the IPQ (Schubert, 2003), the GTS (Yamagishi & Yamagishi, 1994) is a 6-item questionnaire that asses participants' general beliefs regarding the honesty and trustworthiness of others using broad statements on a 5-point scale, ranging from 1 "strongly disagree" to 5 "strongly agree" and the KUSIV3 (Beierlein et al., 2012)(see Chapter 4 for a complete description of the questionnaires). Table 8 summarizes the subjective measurements used in this study. The subjective measurements were implemented with LimeSurvey 4.5.0 (LimeSurvey Development Team, 2012).

5.2.4. Procedere

Figure 20: Experimental Procedure.



Note: LCL = Low Cognitive Load; HCL = High Cognitive Load; VR = Virtual Reality; First, participants provided informed consent, then they watched the introductory video and entered VR to complete the virtual maze either with LCL or HCL, then they completed the intermediate questionnaires about their experience in VR on a laptop (conditions were counterbalanced). Participants then returned to VR and completed one round of the trust game. The experiment concluded with the post-VR measures and the debriefing.

At the beginning of the session, participants read the study information, provided informed consent, and filled in the *pre-VR measures*. After that, the *introductory video* was played where participants were framed that they would interact with a real person during the experiment (for details, see study 2 in Chapter 4 or Lin et al. 2023). *VR tutorial.* After watching the introductory video, participants were equipped with an HMD and VR controllers, then entered the tutorial task to learn how to navigate in VR (for details, see Chapter 4, study 2 or Lin et al. 2023). After the tutorial, participants started with the first VR task escaping from the virtual maze (under either LCL or HCL, depending, on the order of the condition), then the *trust game* (for details, see Chapter 4, study 2 or Lin et al. 2023) and the *intermediate measures*. Upon completion of each CL condition, participants completed the self-constructed intermediate measures. Then returned to VR, where they completed the maze once more under the remaining condition (followed by the second set of intermediate measures) and finally, the *post-VR measures*. Once all tasks were completed under both conditions and the post-VR measures, the experiment concluded with the debriefing and full disclosure of the study's goal. The experimental procedure is summarized in Figure 20.

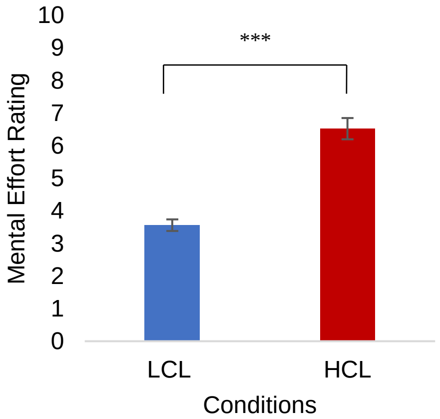
5.2.5. Statistical Analysis

Jamovi version 2.3.21.0 (The Jamovi project, 2021) was used to conduct the analysis. To test for differences in the subjective evaluation of trust towards the VH between conditions, paired sample t-tests were conducted. A mixed ANOVA was conducted to investigate whether the order in which the conditions were presented affected the behavioural proxies of trust. An independent sample t-test was conducted to test differences between the first and second conditions for the main outcome measures (advice seeking, advice following, and response time during the maze task). A Pearson's correlation analysis was conducted to test whether the trust evaluation of the avatar correlates with the number of tokens invested in the trust game. The level of significance was set at $p < .05$.

5.3. Results

Manipulation Check

Figure 21: Manipulation Check



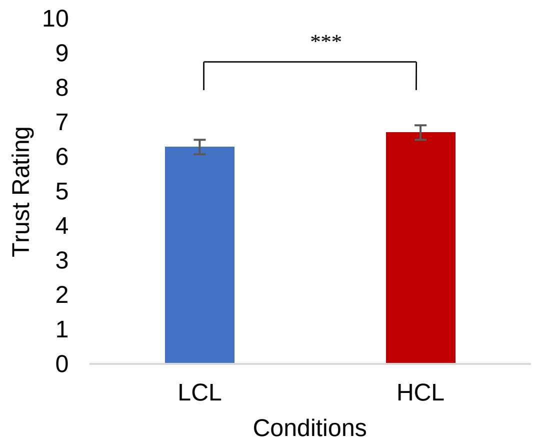
Note: LCL = Low Cognitive Load; HCL = High Cognitive Load; present the significant difference in the experience of mental effort between LCL and HCL condition, the error bars indicate the error bars indicate the standard error, *** $p < .001$.

Due to the non-normal distribution of the data, a non-parametric Mann-Whitney *U* test revealed that higher mental workload was experienced in the HCL condition (*M* = 6.60, *SD* = 2.14) compared to the LCL condition (*M* = 3.56, *SD* = 2.00), *U* (49) = 94.0, *p* <.001 (see Figure 21A). This indicates a successful experimental manipulation. During the secondary task of the HCL condition 102 beeps were played. An additional performance measure revealed that participants on average had 54% "misses" and 46% "hits" of the targeted 32 beeps.

Subjective Trust Evaluation

The avatar was evaluated as significantly more trustworthy after the HCL condition (*M* = 6.70, *SD* = 1.98) compared with the LCL condition (*M* = 6.28, *SD* = 2.14), *t* (49) = 2.29, *p* = 0.026, *d* = 0.32 (see Figure 22). Thus, H1 is rejected.

Figure 22: Virtual Human Trust Evaluation



Behavioural Measures

Advice Seeking

CL had no effect on the number of times advice was asked, *F*(1,48) = 0.08, *p* > 0.774, = 0.002, i.e., the number of times advice was asked did not differ significantly between the LCL condition (M = 10.40, *SD* = 2.53) and the HCL condition (*M* = 10.30, *SD* = 2.55) (see Figure 23A). The order in which CL conditions were presented had no effect on the advice asked, *F* (1, 48) = 0.13, *p* = 0.713, = 0.003. There was no interaction effect between cognitive load and the order of presentation, *F* (1,48) = 0.067, *p* = 0.796, = 0.001. Thus, H2 is rejected.

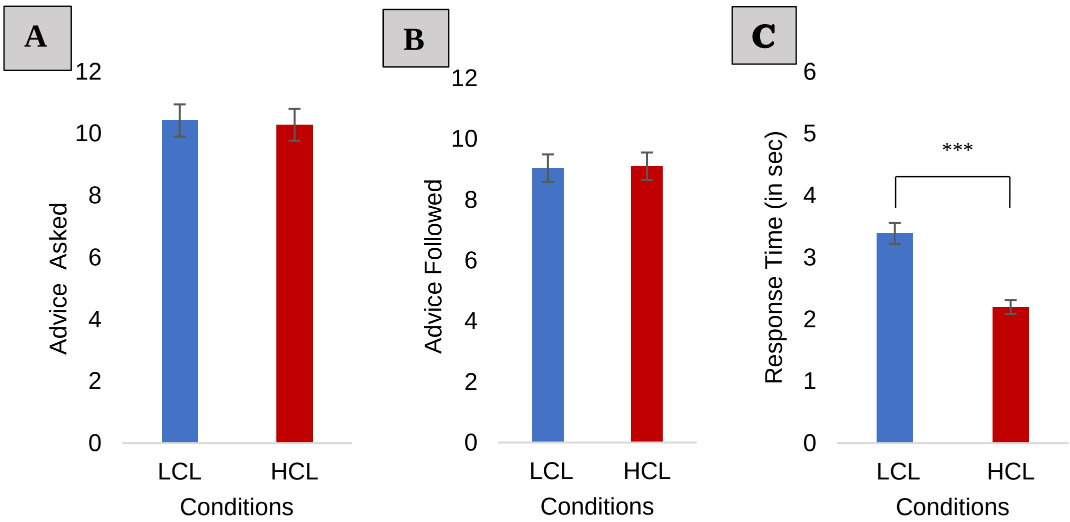
Advice Following

CL had no effect on the number of times advice was followed, F (1,48) = 0.01, p > 0.913, η_p^2= 0.000, i.e., the number of times advice was followed did not differ significantly between the LCL condition (M = 9.04, SD = 2.96) and the HCL condition (M = 9.10, SD = 2.94) (see Figure 23B). The order in which CL conditions were presented had no effect on the advice followed, F (1, 48) = 0.01, p = 0.913, η_p^2= 0.000. There was no interaction effect between cognitive load and the order of presentation, F (1,48) = 0.05, p = 0.828, η_p^2= 0.001. Thus, H3 is rejected.

Response Time to Execute Given Advice

The CL significantly affected participants' response time, F(1,48) = 46.3, p < 0.001, η_p^2 = 0.49. Participants responded to advice 1.52 seconds faster in the HCL (M = 2.20, SD = 1.49) condition compared with the LCL (M = 3.38, SD = 1.22) condition (see Figure 23C). The order in which CL conditions were presented did not affect response time, F(1,48) = 0.33, p = 0.569, η_p^2= 0.01. There was a significant interaction effect between CL and the order in which the conditions were presented in the LCL condition, F (1,48) = 15.1, p < 0.001, η_p^2= 0.24. A post hoc t-test revealed that participants responded to advice significantly faster in the LCL condition, t (48) = -2.68, p = 0.010, d = - 0.757, when the LCL condition was presented as the first condition (M = 2.96 sec, SD = 1.16) compared with when LCL condition was presented second (M = 3.82 sec, SD = 1.13). H4 is accepted.

Figure 23: Behavioural Outcomes Between Conditions



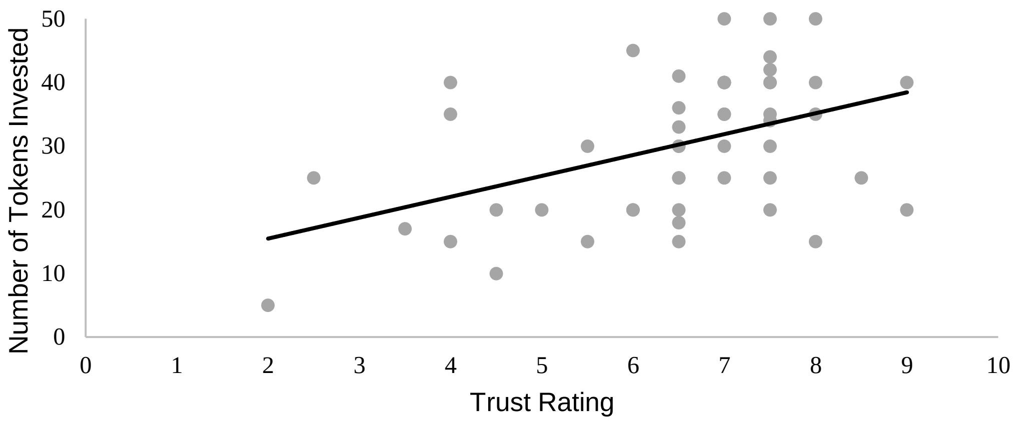
Note: LCL = Low Cognitive Load; HCL = High Cognitive Load; The figure depicts A) the average number of times advice was *asked* and, B) the number of times advice was *followed* in each condition , C) depicts the average *response times*. The error bars indicate the standard error, *** $p < .001$.

Trust Game

Trust Evaluation and Investment of Tokens

The relationship between the number of tokens invested (during the trust game) and the average trust rating of the SIP (as measured explicitly after each condition) was investigated using the Pearson product-moment correlation coefficient. There was a medium, positive correlation between the two variables ($r = 0.5$, $n = 50$, $p < 0.001$**), with a higher explicit trust rating associated with a higher number of tokens invested towards the SIP (see Figure 24). In addition, 44% of the participants felt the avatar would act in their best interest. Thus, H5 is accepted.

Figure 24: Correlation between Tokens Invested and Trust Rating



Note: Scatter plots for the correlation between the number of tokens invested in the trust game and the self-reported trust rating. The line in the plot represents the linear regression between the two variables.

5.4. Discussion

This study investigated the impact of cognitive load on interpersonal trust with a virtual maze task and a trust game under two cognitive load conditions (low cognitive load and high cognitive load). Participants experienced higher mental workload in the high cognitive load condition than in the low cognitive load condition, indicating a successful experimental manipulation. The virtual human was evaluated as significantly more trustworthy after the high cognitive load condition. It could be due to the trustworthy appearance of the virtual human selected for this study (as described in study 2, subsection 4.2.2. in Chapter 4). Despite this, albeit descriptively small, different participants behaved the same during both conditions, they asked and followed advice equally often in the maze. Therefore, the effect of the additional workload may not have been strong enough to influence behavioural decisions of participants. Participants followed advice 9 out of 10 times in both conditions, possibly because trust towards the virtual human was still relatively high in the high cognitive load condition.

In the experimental study of Zhou et al. (2017), which focused on the effects of uncertainty and cognitive load on users' trust in artificial intelligence, it was found that participants are more trusting during uncertainty when sufficient cognitive resources are available to process information. However, when cognitive resources are limited participants have a poorer understanding of the information in the environment, resulting in lower trust toward the social interaction partner. The subjective trust ratings of the present study differ from previous standard non-VR lab studies, which instead observed lower trust evaluations during high cognitive load (Zhou et al., 2017). According to a game theory analysis of repeated games in human-computer interaction, trust is often used as a cognitive shortcut to reduces the complexity of human-computer interactions (Han et al., 2021). It is possible that participants evaluated the virtual human as more trustworthy during the high cognitive load condition as a strategy to reduce the experience of high cognitive load during this condition (Ahmad et al., 2019; Alarcon et al., 2018). According to Ahmad et al. (2019) trust towards a social interaction partner reduce uncertainty and critical investigation of consequences, leading to a decrease in the experience of cognitive load. Furthermore, Khawaji et al. (2014) ascribe the differences in the trust evaluations between cognitive load conditions to the perception that time goes by faster during complex tasks. If participants felt that they escaped faster from the maze during the high cognitive load condition, they could conclude that the advice of the virtual human was helpful. Experimental studies on time perception demonstrated that when attention is allocated to a demanding task, it can become more difficult to accurately process time and perceive time (Brown, 1997; Khan et al., 2006). However, when participants are not aware of time estimation during tasks with less mental effort, the opposite is true (Block et al., 1980, 2010; Block & Gruber, 2014; Hertzum & Holmegaard, 2013). As we did not assess time perception, this explanation remains to be investigated. Although participants trusted the virtual human more during the high cognitive load condition, they behaved the same in both conditions. They were told that the person supposedly controlling the virtual character had a map of the maze, so asking for advice could be perceived as the most obvious choice in both conditions. However, previous research demonstrated that high cognitive load and uncertainty lead to decreased access to mental resources, analytical thinking, and greater reliance on external resources (Ahmad et al., 2019; McBride et al., 2011; Samson & Kostyszyn, 2015). Chen et al. (2016) suggest that participants rely more on external resources during uncertainty even if trust in those resources is low. In this study part of this thesis, participants' behaviour in the maze was not congruent with their subjective evaluation, as no behavioural differences were observed between conditions (they did not ask or follow more or less advice).

Samson and Kostyszon (2015) investigated trust behaviour in a laboratory experiment with the trust game. They found that participant´s behaviour was more impulsive (for example, trusting without careful evaluation or considering the consequences) when their cognitive resources were limited (Samson & Kostyszyn, 2015). Furthermore, individuals may experience time pressure or urgency under a high cognitive load, leading to a bias toward quick and intuitive decisions rather than deliberate and reflective decision-making. This might also explain why participants followed the advice quicker in the high workload condition. In addition, previous experimental research suggests that impulsive decision-making and faster responses or reactions can be expected under high cognitive load (Eggemeier & Wilson, 1991; Mayer & Moreno, 1998; Samson & Kostyszyn, 2015). An experimental study interested in cognitive load and multiplayer games such as The Prisoner‘s Dilemma found that participants under low cognitive load behave more strategically and take more time to make decisions than those under high cognitive load, who are more impulsive and reactive due to their low availability of cognitive resources (Duffy & Smith, 2012).

The order in which the conditions were presented to the participants influenced their response time. During the low cognitive load condition, when the low cognitive load was presented as the first condition, participants responded to advice significantly faster. Participants could be more impulsive and focused on escaping from the maze as quickly as possible instead of thinking about the consequences of trusting/following the given advice. This is in line with the results of Samson and Kostyszyn (2015), who found that participants behave more impulsively when a high cognitive load is experienced. Furthermore, it could also be that the maze task required a lot of cognitive resources to complete for the first time compared with the second time when participants were more familiar with the task and the virtual environment. According to the cognitive load theory, performing a task for the first time requires more cognitive resources than subsequent repetitions (Sweller, 2011). This is because when performing a new task, the brain needs to process and analyze the information, understand the steps or requirements involved, and create new neural connections to encode the task in memory (Sweller, 2011). These processes require cognitive resources such as attention, working memory, problem-solving, and decision-making (Sweller, 2011). The repeated practice leads to experience and familiarity with a task and allows for the development of automaticity, where the task becomes more automated and requires less conscious effort (Sweller, 2011). Therefore, it is also possible that the familiarity with the maze task during the second low cognitive load condition led participants to experience less cognitive load, as more cognitive resources were available to them to think their options through.

5.4.1. Contribution, Limitations, and Future Work

The topic of the current study is highly relevant; however, the study design has some limitations that should be mentioned. A limitation of this study is that we only included a virtual human who is trustworthy in appearance (as described in Study 2, subsection 4.2.2. in Chapter 4). Future research can consider an experimental design with both trustworthy and untrustworthy virtual humans. No feedback on the progress in the maze was provided, and decisions in the maze had no immediate consequences. Although the lack of consequences and feedback was intentional to enhance uncertainty and ambiguity, it probably biased participants towards asking for advice. Furthermore, only female participants were included in this within-subject study, thus, limiting the generalizability of the results.

As social virtual reality gets more popular, investigating how psychological factors influence trust towards virtual characters (agents or avatars) gains importance. For instance, users of social virtual reality can alter their personal information and self-representations leading to privacy and identity issues, such as identity theft and social engineering attacks (Lin & Latoschik, 2022). Knowing that the digital identity of another user could potentially be hacked or stolen can further complicate the trust evaluation process (Falk et al., 2021). Cybercriminals can easily create false identities and manipulate users to act as intended by the attacker (Falk et al., 2021; Montañez et al., 2020). The results from this study demonstrate that participants who experience high cognitive load might be vulnerable to making impulsive decisions.

5.4.2. Conclusion

The results demonstrated higher trust towards the avatar after the high cognitive load condition, albeit with a small effect size that did not have any behavioural consequences for advice asking and following in the novel maze paradigm designed to investigate trust in virtual reality. Importantly, participants followed advice quicker in the high cognitive load condition, suggesting that more impulsive decisions were made due to fewer cognitive resources.

Considering the rising importance of social virtual reality, identifying and investigating factors such as cognitive workload can help to enhance user experience and improve the security of digital interactions. However, further research on trust toward virtual humans and the psychological factors influencing trust is needed.

6. Study 4: How we feel

The Impact of Emotional Affect on Trust Towards a Virtual Human in Virtual Reality

6.1. Introduction

Emotions are crucial in everyday trust decisions, as emotions influence our perception of trustworthiness, our interpretation of behaviour, and our ability to make rational decisions (Lee & Selart, 2012). If our emotional state is negative, for example, feeling stressed or anxious, we may be more likely to be suspicious of others or make decisions based on fear rather than objective evidence (Jung et al., 2014; Lee & Selart, 2012). The experimental work of Dunn & Schweitzer (2005) induced positive and negative emotional affect in participants by asking them to describe a past event (either positive or negative). After that, participants were tasked to rate the trustworthiness of an unfamiliar co-worker. The results showed that participants in the positive condition were more trusting (evaluated the co-worker as more trustworthy) than those in the negative condition (Dunn & Schweitzer, 2005).

Another series of experiments includes emotional affect (EA) manipulations that investigate how emotions influence logical reasoning (Jung et al., 2014). The participant's emotional state was altered by providing feedback on their performance on an intelligence task. After that, they had to complete a set of logical inference problems. The results showed that emotions impact reasoning performance, as participants in the negative emotional state had poorer logical reasoning than those in the positive emotional state (Jung et al., 2014). The results are important when considering how a negative emotional state influences logical reasoning, decision-making, and daily problem-solving (Jung et al., 2014). Furthermore, Myers & Tingley (2016) investigated how emotions influence trust in an experimental study using the trust game and found that emotions can significantly impact trust decision-making in social interactions. In their results, they look beyond the valence-based approach of emotions and conclude that negative emotions can decrease trust, but only if those negative emotions (such as anxiety and anger) produce doubt or lack of confidence in information (low certainty appraisals) (Myers & Tingley, 2016). The findings of this study in this thesis are explained by the study of Meyers and Tingley, (2016). The "Affect-as-Information" theory, a concept within the realm of psychology, suggests that individuals often rely on their emotional states as valuable sources of information when making judgments or decisions. Essentially, people tend to use their current emotional experiences as shortcuts or cues to evaluate situations, even if they are not consciously aware of the specific reasons behind those emotions (Schwarz, 2012). For instance, if someone is feeling uneasy or anxious while making a decision, they might use these emotional signals as a way to assess whether the situation is risky or unsafe, without necessarily knowing all the underlying details. In cases where the source of the emotion is not clear or easily identifiable, individuals may inadvertently attribute their feelings to the situation at hand, even if there might be other external factors contributing to their emotions (Schwarz, 2012). Therefore, once individuals are informed about the origin of their emotions, the impact of those emotions on evaluations of trust is neutralized as the emotions cease to provide meaningful information (Lerner & Keltner, 2000). Additionally, according to an experimental study investigating trust in highly automated driving, anxious and uncomfortable participants are less inclined to trust the environment or others in the environment (Kraus et al., 2020). Furthermore, individuals with depression often struggle with interpersonal trust (Kim et al., 2012; Lester & Gatto, 1990).

On the other hand, happiness is positively correlated with trust and is a consequence of prosocial behaviour (Lane, 2017). A study conducted by Jasielska (2020) involved student participants who participated in a trust game and subsequently completed assessments of trust, happiness, and kindness. The findings indicate that kindness enhances the relationship between trust and happiness. Furthermore, happy individuals are more inclined to trust when they engage in prosocial activities (Jasielska, 2020). The connection between these insights becomes clearer when considering how trust and its correlation with happiness align with the mediation analysis findings. Mediation analysis suggests that the influence of positive and negative valence in emotions is fully mediated by the perception of a lack- of - control over emotions (Dunn & Schweitzer, 2005). If people feel and perceive to be in control of their emotional experience, then their decisions can change (Dunn & Schweitzer, 2005). Additionally, experimental laboratory studies found that emotions such as empathy and rapport can build trust between people and increase collaboration and communication (Argelaguet et al., 2016; Menges et al., 2022).

Several social scientists have used VR technology to manipulate emotions or investigate emotions through VR environments and experiences (Felnhofer et al., 2015; Radiah et al., 2023; Sansoni et al., 2022; Susindar et al., 2019). For example, Felnhofer et al. (2015) designed five emotion-inducing virtual park scenarios to induce specific emotional affective states and found VR effective. A user study investigating the influence of avatar personalization on emotions in VR found significant differences in happiness with the personalized same-gender avatar and the personalized opposite-gender avatar. Researchers use VR to create an innovative training experience in VR that helps cancer patients overcome negative emotions and physical realities related to their diagnosis (Sansoni et al., 2022). Furthermore, an experimental study compared the effectiveness of emotional induction between a desktop computer and VR. The results show the effectiveness of using VR as a method for inducting emotions when studying decision-making under various emotional conditions (Susindar et al., 2019).

As described above, VR is an effective tool to induce emotions, create ecologically valid environments to help patients process emotions and control these environments to study emotions. There is however little to no research on how emotional affect (EA) influences trust towards virtual humans in virtual reality. Therefore, in the present study, we investigate whether manipulating EA influences the evaluation of a SIP's trustworthiness and the consequent trust behaviours. In study 2 (see Chapter 3) we modified and validated a variant of the virtual maze paradigm of Hale et al. (2018) that measures interpersonal trust towards VHs as SIPs in VR (also see Lin et al. 2023). The present study used this paradigm to investigate the impact of EA on the trust evaluation of VHs and the consequent trust behaviours. EA was manipulated with the written version of the "Autobiographical Emotional Memory Task" (AEMT) as the primary manipulation. Throughout the experimental tasks, participants were exposed to either positive or negative pictures and sounds to strengthen the primary manipulation resulting in either a positive or negative emotional affect condition. We hypothesized that negative emotional affect would lead to lower trust scores and that fewer trust behaviours would be observed compared with the positive emotional affect condition, where higher trust scores and more trust behaviours are expected, based on previous research (Dunn & Schweitzer, 2005; Kugler et al., 2020; Tislar et al., 2014).

6.1.1. Hypotheses

The reviewed literature inspired the following hypotheses:

H1: Participants evaluate the SIP as less trustworthy in the negative emotional affect condition (NEA).

H2: Participants show less advice-seeking behaviour in the NEA condition.

H3: Participants display less advice-following behaviour in the NEA condition.

H4: Participants respond to advice slower in the NEA condition.

For the trust game, we expected the following:

H5: Participants who evaluate the VH as trustworthy invest more tokens.

H6: Participants invest less tokens in the NEA condition.

6.2. Methods and Materials

6.2.1. Study Design and Participants

In order to reveal at least a medium effect (*d* = 0.5, β = 0.90, α = 0.05), a calculation with G*power (Faul et al., 2007) indicated that 44 participants would be required for this. In total, 50 healthy women (*M* = 23.3 years, *SD* = 4.10, range: 18 and 35) with no psychiatric or neurological disorders were recruited via the university's participant recruitment system (SONA). Every participant was compensated with 15 euros for participating in the experiment that lasted 1.5 hours. The impact of emotional affect was investigated with a within-group design, and each participant took part in a positive emotional affect condition (PEA) and a negative emotional affect condition (NEA). The emotional affect was the within-subject factor, and the order of the conditions was counterbalanced to ensure no order effects. The Ethical Committee of the Psychological Institute of the Faculty of Human Sciences of the University of Würzburg reviewed the protocol and provided ethical clearance (GZEK 2022-65).

6.2.2. Apparatus

The experiment was implemented as an immersive VR experience using Engine, 2020 Unity.3.14f and ran on a VR-capable PC (Intel Core i7-2600K, Nvidia GTX 970 16GB RAM). The VR hardware consisted of an Oculus Quest 2, HMD, and two controllers connected to the PC through an Oculus Link cable (or wireless). The virtual environments for the virtual maze task, the trust game, and the trustworthy VH were the same as those in study 2 (validation study) and as described in Lin et al. 2023 and Chapter 4.

6.2.3. Emotional Affect Manipulation and Measurements

Manipulation of Emotional Affect

The experimental conditions differed in the EA (either positive or negative, depending on the condition). The *primary manipulation* of EA relied on the widely used "Autobiographical Emotional Memory Task" (AEMT; Mills & D'Mello, 2014) to induce either positive or negative EA (depending on the condition) (see Appendix D: A and B). Participants were asked to complete this task on paper with a pen with a 10-minute time limit. They were informed that this task would not be analysed and that it was an essential part of the experimental process. Participants took the sheet home with them after the experiment. The confidential nature of this task encouraged participants to connect to their emotions and to be as honest as possible while recalling emotional events.

During the *NEA* condition, participants were asked to recall a negative situation that made them feel out of control, anxious, or frustrated (see Appendix D: A). During the *PEA* condition, participants were asked to recall a positive situation that made them happy and/or joyful. It can either be in the past or something they are experiencing at present (see Appendix D: B).

While participants were immersed in VR, they were exposed to a total of 12 pictures from the "International Affective Picture System" (IAPS; Lang et al. 2008) and 12 sounds from the "Affective Auditory Stimulus Database" (IADS-E; Yang, 2018) that were presented simultaneously to *strengthen* the initial manipulation (one picture and one sound at a time). Comprehensive information regarding the normative ratings of pictures and sounds on the SAM scale (Bradley & Lang, 1994) can be found in the appendix (for the NEA condition, Appendix E, and for the PEA condition Appendix F) (see Table 9 for a summary). In the PEA condition pictures, and sounds were selected based on high valence and dominance, and in the NEA condition, based on low valence and dominance.

Table 9: Pictures and Sounds on the Self - Assessment Manikin Scale (SAM)

SAM	Negative (NEA)				Positive (PEA)			
	Pictures		Sounds		Pictures		Sounds	
	M	SD	M	SD	M	SD	M	SD
Valence	1.55	1.09	3.69	1.36	8.17	0.09	7.56	0.44
Arousal	6.58	0.70	5.14	1.06	4.83	0.65	7.26	1.04
Dominance	2.94	0.91	2.13	0.34	6.73	0.62	5.78	0.57

Note: NEA = Negative Emotional Affect; PEA = Positive Emotional Affect; SAM = The Self-Assessment Manikin Scale; VR = Virtual Reality; Summary of the 12 pictures and sounds used in the PEA and NEA condition, rated on the SAM scale from 1-9 by participants see appendix E and F for details (*IADS-E*, n.d.; *International Affective Picture System (IAPs) | Research Administration*, n.d.).

Figure 25: Emotional Affect Manipulation and SAM scale in Virtual Reality

Note: SAM = The Self-Assessment Manikin Scale; A) Trigger warning in VR that negative content may be presented. B) Example of a positive picture with a high vividness and the SAM scale. C) Example of a negative picture with decreased vividness and the SAM scale. Permission was granted to use both picture and sound databases.

The pictures and sounds were displayed for 6 seconds and then faded out for 6 seconds. Furthermore, before the trust game, participants were exposed to one picture and one sound (positive or negative, depending on the condition) before investing tokens. Participants were made aware of the type of pictures and sounds they could expect during the study before the recruitment process. A trigger warning was implemented before the maze task in VR (see Figure 25A) to warn participants of exposure to potentially disturbing images and sounds (in both conditions).

Measurement of Emotional Affect

Manipulation Check

We measured the EA of participants after the AEMT (1 and 2) with the German version of the "Positive and Negative Affect Schedule" (PANAS: Breyer & Bluemke, 2016). The German version of the PANAS was adapted from the widely used English-language emotional state assessment instrument PANAS which was created by Watson et al. (1988). The questionnaire comprises 20 items encompassing various sensations and feelings. Out of these, 10 items pertain to dimensions of PEA while the remaining 10 items focus on dimensions of NEA. Participants had to indicate to what degree the 20 words described their current emotional state on a 5-point scale ranging from (1) not at all to (5) extremely. Further manipulation checks were conducted during the intermediate measures after each task, as described below.

Questionnaires

The *pre-VR measures* include the participant's report of demographic and general health information and the BFI personality assessment (Rammstedt & John, 2007) (see Chapter 3 for a complete description of the BFI). Emotional valence, arousal, and dominance were assessed during VR exposure with the "Self-Assessment Manikin" (SAM; Bradley & Lang, 1994) while participants were exposed to the pictures and sounds. The SAM is a technique that uses non-verbal images to directly measure the emotional response of individuals to various stimuli, assessing their levels of pleasure, arousal, and dominance. In SAM, the pleasure dimension is represented by a spectrum ranging from a smiling, happy figure to a frowning, unhappy figure. The arousal dimension is depicted through a range from an excited, wide-eyed figure to a relaxed, sleepy figure. The dominance dimension is conveyed by changes in the size of SAM, with a larger figure indicating maximum control in the given situation. In the digital version implemented for this study, the participants used sliders within the virtual environment to mark an 'x' over any of the five figures within each scale or between two figures, effectively creating a 9-point rating scale for each dimension (see Figure 25B and C for the digital implementation in this study and Figure 26 for an example of the figures on the SAM scale).

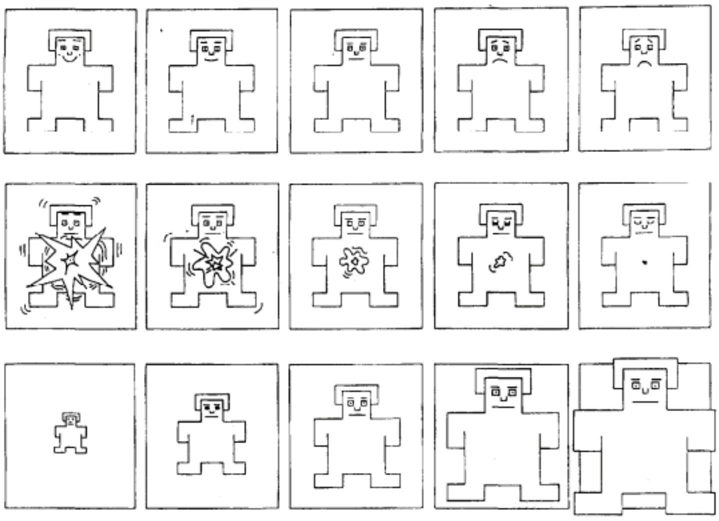
Table 10: Descriptive Questionnaire Data

	Pre – VR Measures	
	M	**SD**
Age	23.32	4.10
BFI Extraversion	3.48	1.00
BFI Agreeableness	3.57	0.81
BFI Conscientiousness	3.27	0.44
BFI Neuroticism	2.88	0.51
BFI Openness	3.39	0.94
	Post – VR Measures	
KUSIV3	3.41	0.45
STAI Trait	1.77	0.46
SSQ Total	12.02	13.74
SSQ Nausea	26.68	19.08
SSQ Oculomotor	29.51	25.87
SSQ Disorientation	17.28	13.11
IPQ Spatial Presence	2.90	0.91
IPQ Involvement	2.18	0.83
IPQ Experienced Realism	1.48	0.64
IPQ General	4.26	1.56

Note: BFI = Big Five Inventory; KUSIV3 = Shortened German Version of the Interpersonal Trust Scale of Rotter; STAI Trait= Trait Anxiety Inventory; SSQ = Simulator Sickness Questionnaire; IPQ = Igroup Presence Questionnaire.

During the *intermediate measures*, participants had to report their experience of the tasks on a 10-point scale (ranging from 1- *negative* to 10 - *positive*) (manipulation check). For example, "How would you describe your emotional state during the maze." After the task, participants reported their trust towards the VH on a 10-point scale (ranging from 1- *not at all* to 10 - *completely*). The *post-VR measures* included the SSQ (Kennedy et al., 1993) and the IPQ (Schubert, 2003), KUSIV3 (Beierlein et al., 2012) (see Chapter 4 for a complete description of the SSQ, IPQ, and KUSIV3), and the STAI trait (Laux, 1981) (see Chapter 3 for a full description of STAI). The subjective measurements were implemented with LimeSurvey 4.5.0 (LimeSurvey Development Team, 2012). Table 10 summarizes the means and standard deviations for the subjective measurements used in this study.

Figure 26: The Self - Assessment Manikin Scale (SAM)



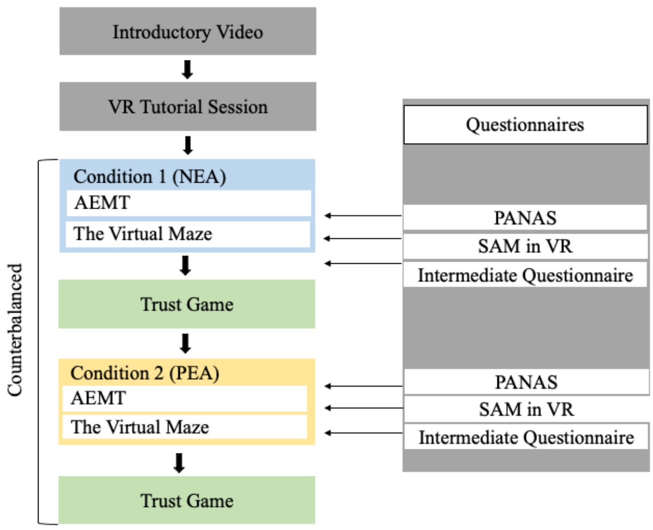
Note: The Self-Assessment Manikin (SAM) utilized for rating the affective dimensions of valence (top panel), arousal (middle panel), and dominance (bottom panel).

Main Outcome Measures

The behavioural measurements (main outcome measure) were the same as in the validation (in Chapter 4, study 2 or Lin et al. 2023) and cognitive load study (in Chapter 5, study 3) (i.e., 1) the number of times advice was asked and 2) the number of times participants followed the advice (if applicable), 3) the response time to execute the advice (if applicable), 4), the number of tokens invested in a trust game.

6.2.4. Procedure

Figure 27: Experimental Procedure



Note: NEA = Negative Emotional Affect; PEA = Positive Emotional Affect; AEMT = Autobiographical Emotional Memory Task, PANAS = Positive and Negative Affect Schedule, SAM = The Self-Assessment Manikin Scale; VR = Virtual Reality; all participants experience all conditions – order of conditions was counterbalanced.

At the beginning of the session, participants read the study information, provided informed consent, and filled in the *pre-VR measures*. After that, the *introductory video* was shown to convince the participants that they would interact with a real person during the experiment (for details see Chapter 4, study 2 or Lin et al. 2023). *VR tutorial.* After watching the introductory video, participants were equipped with an HMD and VR controllers, then entered the tutorial task to learn how to navigate the VR (for details, see study 2 in Chapter 4 or Lin et al. 2023). For half of the participants, the experiment started in the NEA condition with the written AEMT task, and the first PANAS was completed, then the *maze task* and *trust game* followed (for details, see Chapter 4, study 2 or Lin et al. 2023). After that, they completed the *intermediate measures*, followed by the *first part of the debriefing,* where they watched a short (3 minutes) funny video of dogs climbing stairs and read 30 positive Velten statements (slide show video of 5 min and 30 seconds) (see Appendix E). Hereafter they proceeded to the PEA condition. Participants who completed the PEA condition skipped the first part of the debriefing and immediately completed the intermediate measures after the trust game. Finally, *post-VR measures* were completed once all tasks were completed under both conditions, followed by the *second part of the debriefing*. Participants also received a list of contact details of crisis centers contact in case the experiment triggered past trauma (Appendix F) (see Figure 27 for a summary of the experimental procedure).

6.2.5. Statistical Analysis

Jamovi version 2.3.21.0 (The Jamovi project, 2021) was used to conduct the analysis. A paired sample t-test was conducted to test differences in PANAS and SAM ratings between conditions. To test for differences in the subjective evaluation of trust towards the VH between conditions, paired sample t-tests were conducted. A mixed ANOVA was conducted to investigate whether the order in which the conditions were presented affected the behavioural proxies of trust. An independent sample t-test was conducted to test differences between the first and second conditions for the main outcome measures (advice seeking, advice following, and response time during the maze task). A Pearson's correlation analysis was conducted to test whether the trust evaluation of the avatar correlates with the number of tokens invested in the trust game. The level of significance was set at $p < .05$.

6.3. Results

Manipulation Check

Table 11: Manipulation Check

	NEA		PEA					
	M	**SD**	**M**	**SD**	**t**	***p***	***d***	**95% CI**
PANAS	8.60	11.09	21.02	7.21	-7.90	<.001*	1.12	-15.6, - 9,26
SAM Scale								
Valence	3.48	1.78	7.05	1.39	-41.90	<.001*	-1.71	- 3.75, - 3.40
Arousal	5.56	1.68	4.37	1.90	15.40	<.001*	0.63	1.04, 1.35
Dominance	4.28	1.73	5.60	1.83	-16.20	<.001*	-0.66	- 1.48, - 1.16
Subjective experience of emotional affect during tasks								
Maze Task	3.16	1.56	7.58	1.85	-17.2	<.001*	- 2.43	- 4.94, - 3.90
Trust Game	4.62	1.63	6.90	2.03	7.75	<.001*	- 1.1	- 2.87, - 1.69

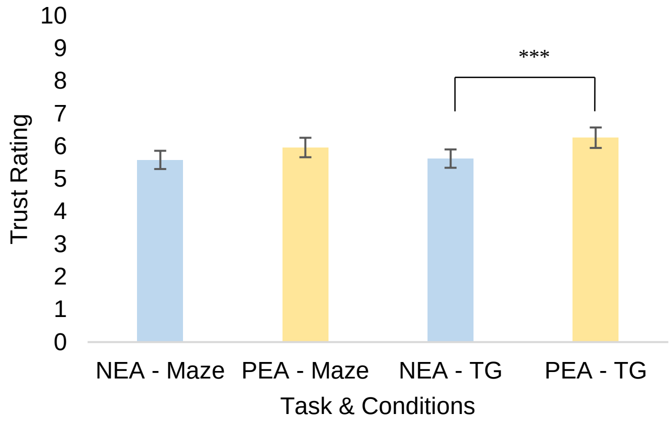
Notes: NEA = Negative Emotional Affect; PEA = Positive Emotional Affect; PANAS = Positive and Negative Affect Schedule, SAM = The Self-Assessment Manikin Scale; The table shows the significance difference between the means of the PANAS and SAM between conditions, *** $p < .001$.

To test the success of the manipulation of EA, a series of paired sample t-tests was conducted (see Table 11). A paired sample t-test indicated that participants experienced a significantly more negative emotional state on the PANAS after the primary manipulation (AEMT) in the NEA condition. Furthermore, the in-the-moment manipulations in VR indicated that participants experienced significantly higher arousal in the NEA condition and significantly more positive valence and higher dominance in the PEA condition on the SAM scales. Participant's subjective reports of their EA after completing the experimental tasks (Maze Task and Trust Game) indicated a significantly more negative emotional affect in the NEA condition. Hence, the manipulations were successful.

Subjective Trust Evaluation

There was no significant difference in the participant's trust evaluation of the VH in the maze task between the NEA condition (M = 5.58, SD = 2.00) and the PEA condition (M = 5.96, SD = 2.17), t (49) = 1.55, p = 0.128, d = 0.22 (see Figure 28). However, in accordance with the first hypothesis, participants evaluated the VH as significantly more trustworthy in the PEA condition (M = 6.26, SD = 2.11) of the trust game compared to the NEA condition (M = 5.62, SD = 1.93), t (49) = 2.23, p = 0.030. d = 0.32. Thus, H1 is partially accepted.

Figure 28: Virtual Human Trust Evaluation



Note: VH = Virtual Human; TG = Trust Game; NEA = Negative Emotional Affect; PEA = Positive Emotional Affect; A summary of the virtual human's trust evaluation between conditions in the maze task and trust game. The error bars indicate the standard error, *** p < .001.

Behavioural Measures

Advice Seeking

EA had no significant effect on the number of times advice was asked, F (1,48) = 0.35, p = 0.555, η_p^2 = 0.007, i.e., the number of times advice was asked did not differ significantly between the PEA condition (M = 7.46, SD = 3.24) and the NEA (M = 7.26, SD = 2.88) condition (see Figure 29A). The order in which EA conditions were presented had no effect on the advice asked, F (1,48) = 1.58, p = 0.216, η_p^2 = 0.032. However, there was a significant interaction effect between the EA and the order of the conditions presented in the NEA condition, F (1,48) = 21.45, p < 0.001, η_p^2 = 0.309. A post hoc t-test revealed that participants asked significantly less advice, t (48) = -3.41, p < 0.001, d = -0.964 when the NEA condition was presented as the first condition (M = 6.00, SD = 3.01) compared with when the NEA condition was presented second (M = 8.52 sec, SD = 2.14). Thus, H2 is rejected.

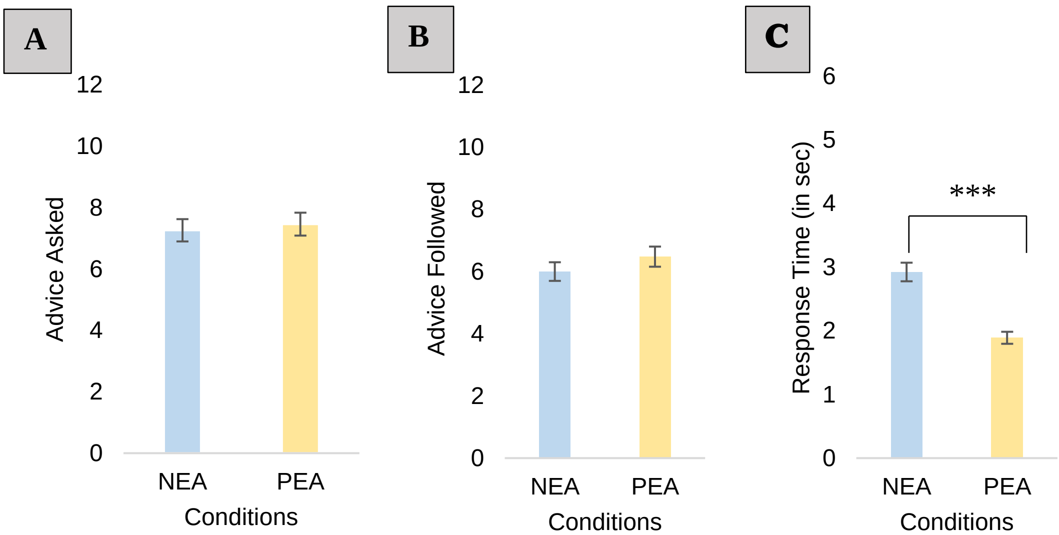
Advice Following

EA had no effect on the number of times advice was followed, F (1,49) = 2.54, p = 0.117, η_p^2 = 0.050, i.e., the number of times advice was followed did not differ significantly between the PEA condition (M = 6.48, SD = 3.17) and the NEA condition (M = 6.00, SD = 2.74) (see Figure 29B). The order in which EA conditions were presented did not affect the advice followed, F (1,48) = 0.33, p = 0.568, η_p^2 = 0.007). However, there was a significant interaction effect between the EA and the order of the conditions presented in the NEA condition, F (1,48) = 19.24, p < 0.001, η_p^2 = 0.286. A post hoc t-test revealed that participants followed significantly less advice, t (48) = -2.38, p = 0.022, d = -0.672 when the NEA condition was presented as the first condition (M = 5.12, SD = 3.10) compared with when the NEA condition was presented second (M = 6.88, SD = 2.03). Thus, H3 is rejected.

Response Time to Execute Given Advice

The EA significantly affected participants' response time, F (1,48) = 30.69, p <.001, η_p^2 = 0.390, i.e., participants responded to advice significantly slower in the NEA condition (M = 2.92, SD = 1.42) than in the PEA condition (M = 1.89, SD = 1.11) (see Figure 29C). The order in which EA conditions were presented had no effect on response time, F (1,48) = 1.08, p = 0.303, η_p^2 = 0.022. However, the order in which EA conditions were presented between participants had a significant effect on response time, F (1,48) = 4.46, p = 0.040, η_p^2 = 0.085. There was a significant difference in the response time between the first (M = 2.24, SD = 1.52) condition and the second (M = 2.33, SD = 1.21) condition, independent of the emotional affect condition, t (98) = 0.704, p = 0.482, d = 0.141. H4 is accepted.

Figure 29: Main Outcome Measures



Note: NEA = Negative Emotional Affect; PEA = Positive Emotional Affect. The error bars indicate the standard error, *** p < .001.

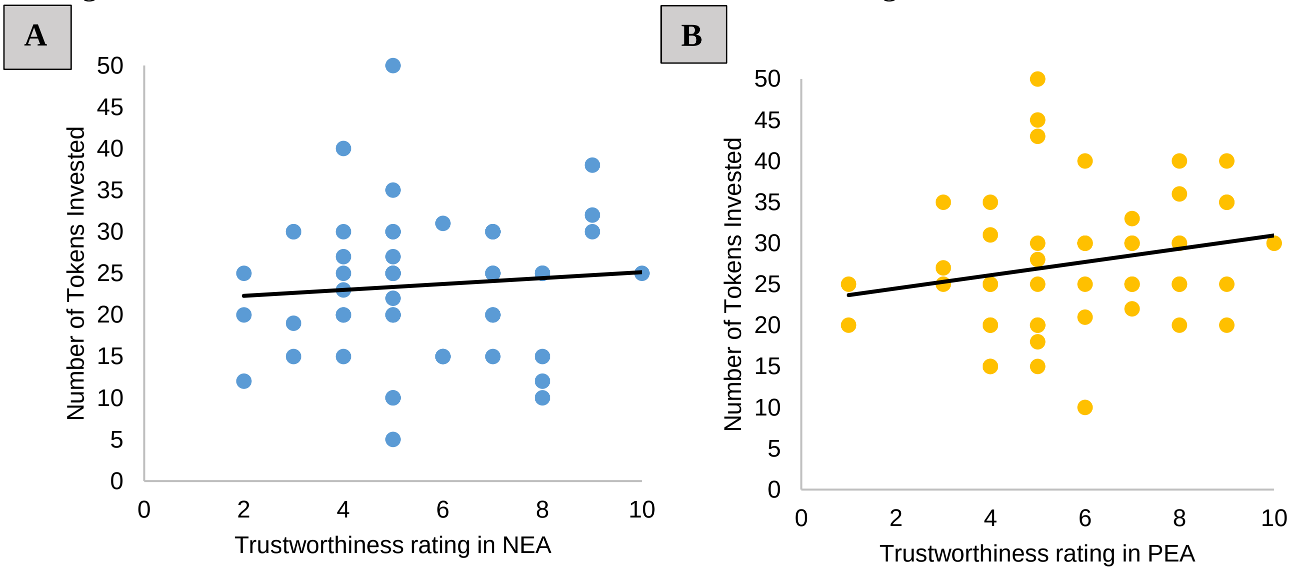
Trust Game

Trust Evaluation and Investment of Tokens.

There was a significant positive correlation between the number of tokens invested (during the PEA trust game) and the trust ratings towards the VH in the PEA condition (r =

0.287, n = 100, *p* = 0.002) (see Figure 30). There was no correlation between the number of tokens invested in the NEA condition and the trust ratings towards the VH in this condition. Thus, H5 is partially accepted. Participants invested significantly more tokens in the PEA (*M* = 26.9, *SD* = 9.34) condition compared to the NEA (*M* = 24.4, *SD* = 7.67) condition (*t* (49) = 2.14, *p* = 0.037, *d* = 0.303. H6 is accepted.

Figure 30: Correlation between Tokens Invested and Trust Rating.



Note: NEA = Negative Emotional Affect; PEA = Positive Emotional Affect; Scatter plots for the correlation between the number of tokens invested in the trust game and the self-reported trustworthiness rating of the VH in the NEA and PEA condition. The lines in the plots represent the linear regression between the two variables.

6.4. Discussion

This study investigated the impact of emotional affect on trust toward a virtual human in a virtual maze task and a trust game under two emotional affect conditions (negative emotional affect vs. positive emotional affect). Participants experienced a more negative emotional affect in the negative condition compared to the emotional affect experienced in the positive condition, indicating a successful experimental manipulation. Participants evaluated the virtual human equally as trustworthy in both conditions in the maze task. A possible reason for this is that the appearance of the virtual human selected for this study was evaluated as trustworthy and medium in attractiveness. Furthermore, the virtual human's behaviour was consistent in both conditions (friendly and trustworthy) (Lin et al., 2023). According to the mere-exposure theory (Zajonc, 1968), it could be that repeated exposure to one trustworthy virtual human in both conditions reduced the participant's experience of uncertainty and increased comfort and familiarity, which strengthened the evaluation of the virtual human's trustworthiness. The number of times advice was asked or followed was similar between the conditions. It is possible that the negative emotional affect manipulation did not induce enough uncertainty in participants to influence their trust and behaviour towards the virtual human. An experimental study that investigated the effect of emotion on trust in a dyadic relationship found that negative emotional states can decrease trust, but only if the negative emotional state leads to *high* degrees of uncertainty (Myers & Tingley, 2016).

Furthermore, during the maze, the most accessible option for participants was to ask for advice. In the positive emotional affect condition, it might have been possible that participants sought advice due to the enhanced cognitive mechanisms associated with positive emotions facilitating approach and trust behaviours, such as seeking and following advice (Bradley et al., 2001; Cacioppo et al., 1999; Davidson et al., 2007). When participants asked for advice in the negative emotional affect condition, it could be an attempt to change their emotional state by relying impulsively on external resources, according to previous experimental studies (Biros et al., 2004; Eben et al., 2020; Ruff et al., 2002). When the negative emotional affect condition was presented as the first condition participants asked and followed significantly less advice. This could be due to the initial evaluation of the virtual human as a stranger, as they interacted with it for the first time, they might have been more hesitant that the virtual human would help them (Sandstrom et al., 2022). An experimental study investigating psychological barriers to social connection with strangers found that people are more pessimistic and hesitant to interact with strangers (Sandstrom et al., 2022).

In contrast, prior interaction with the virtual human (in the negative emotional affect condition as the second condition), participants could have experienced a sense of familiarity with the virtual human due to the repeated interaction (Cochard et al., 2004; Sandstrom et al., 2022). Furthermore, participants could also be less confident and self-assured when the negative emotional affect condition was presented as the first condition due to unfamiliarity with the environment, the task, or the virtual human. An experimental study investigating the emotional impact of uncertainty found that high degrees of uncertainty increased the intensity of negative emotional states (Morriss et al., 2022). Therefore, the negative emotional manipulation further strengthened participants' uncertainty and self-assurance, leading to fewer trusting behaviours (asking and following advice) (Morriss et al., 2022). If the negative emotional affect condition were presented second, participants would have been more familiar with the environment, the task, and the virtual human.

Furthermore, the virtual human was evaluated as equally trustworthy in both conditions, and the participant's behaviour was due to their trust in the virtual human and not because of their emotional affect. Importantly, when participants executed the advice, they responded to advice slower in the negative emotional affect condition. In an experimental study, researchers used electroencephalography technology with a cued-action task and found that negative emotion decreases movement speed (Li et al., 2019). Previous studies show that emotional affect influences the motor system in the brain to move and respond slower, as more cognitive resources are used for emotional processing (Jung et al., 2014; Krasovsky, 2022; Li et al., 2019). Thus, it might be possible that in the negative emotional affect condition, participants used more cognitive resources to process their emotional state, leading to increased caution and hesitancy in their actions. As a result, they exhibited delayed response times in the virtual maze task.

In the trust game, participants evaluated the virtual human as more trustworthy in the positive emotional affect condition. It could be that participants were more certain and confident in the positive emotional affect condition because trust was built through repeated interaction with the virtual human during the virtual maze (as suggested by Zajonc, 1968). Furthermore, the positive emotional affect leads to prosocial behaviour and confidence in participants (Caprara et al., 2022; Mesurado et al., 2021). It could also be that the trust game is a more sensitive instrument to measure trust compared to the virtual maze task. In line with the empirical work of Leahy & Whited (1996), it shows that high levels of certainty lead to more investments and higher trust in the market. While positive emotions create a sense of familiarity, comfort, and a more accurate assessment of trustworthiness (Bradley et al., 2001; Lount, 2010; Serva et al., 2005; Spering et al., 2005), negative emotions can create suspicion and mistrust. This leads to trust being damaged during the negative emotional affect condition in the virtual maze and results in lower trust and fewer investments towards the virtual human in the trust game. Thus, we can conclude that trust was built during the positive emotional affect condition of the maze task, and the effect is seen in the trust game, as participants invested more tokens in the positive emotional affect condition and rated the virtual human as significantly more trustworthy in the trust game. The underlying assumption is that the higher the trust towards the virtual human, the more tokens participants invested.

Consequently, this finding is congruent with other experimental studies using the trust game (Alós-Ferrer & Farolfi, 2019; Bejarano et al., 2021; Cochard et al., 2004). To summarize, the present experimental study found no effect of emotional affect on trust, firstly because the emotional manipulation was not strong enough and secondly because the virtual human (appearance/behaviour) was evaluated as trustworthy in both conditions of the virtual maze. Thus, asking, and following advice was the participant's best option. Trust was built during the positive emotional affect condition and damaged during the negative emotional affect condition of the virtual maze. This effect is seen in the participant's investment behaviour during the trust game, with more tokens invested during the positive emotional affection condition of the trust game.

6.4.1. Contribution, Limitations, and Future Work

The advancements in social virtual reality come with new risks of abusing human trust and emotions (Biegelman, 2013). Investigating the impact of emotional affect on trust in virtual reality is important as it influences the user experience, trust-building processes, behavioural responses, application effectiveness, and ethical considerations. By understanding and considering these dynamics, designers, and developers can create more immersive, engaging, and trustworthy virtual reality experiences. Trust is fundamental in human-computer interactions, including virtual reality (Chiou et al., 2020). It is important that users trust the technology and that it must be designed or implemented in a way that conveys its trustworthiness (Lin & Latoschik, 2022). Emotional affect can significantly influence the development of trust (Jung et al., 2014). When users feel positive emotions in virtual reality, they are more likely to trust the virtual environment, system, and information. Trust is crucial for users to feel safe, engaged, and willing to interact with the virtual world. Designers and developers must consider the ethical implications of manipulating users' emotions (Kim et al., 2020; Rebelo et al., 2012). Understanding the impact of emotional affect on trust helps ensure that users are not manipulated or harmed psychologically (Danaher & Sætra, 2022). It allows for developing responsible and ethical guidelines for creating emotionally engaging virtual reality experiences (Maalem Lahcen et al., 2020). For example, understanding how emotional affect influence trust in virtual social interactions can prevent the misuse of emotions from persuading their victims to take an action they would not usually take, such as sharing their personal information with a nefarious actor (Mouton et al., 2016; Wang et al., 2021). Opposite effects are also true possible depending on how the information is being used and which intentions programmers have.

This study has limitations: the impact of the colours of the doors was not assessed at the validation stage of the virtual maze paradigm. It could be that participants in this study made decisions based on the door colour preference, or the colours of the doors could have had an emotional impact on participants, which was not assessed (Kaur, 2020). Another limitation of the study design could be that a within-study design was used and only a trustworthy virtual human was included in this study. It would be interesting to include one trustworthy and one untrustworthy virtual human in a future study to investigate the differences in trust evaluation under different affective states. Understanding how emotional affect influences trust is crucial as we continue to develop and integrate virtual reality into various aspects of our lives. This understanding can support developers in creating more immersive and engaging experiences and inform the design of training programs to improve communication and collaboration skills in virtual environments. Furthermore, it can assist with the prevention of social engineering attacks that exploit trust and human emotions.

6.4.2. Conclusion

The results in the trust game demonstrated that trust was built during the positive maze condition, as participants evaluated the virtual human as more trustworthy (compared to the negative condition of the trust game), and as a result, participants invested more tokens. During the virtual maze task, participants evaluated the virtual human's trustworthiness equally as high under both conditions, and they also behaved the same. Importantly, participants followed advice slower in the negative emotional affect condition, suggesting that more cognitive resources were used to process their emotional state. This study aims to serve as a catalyst for future investigations into the dynamics of trust in virtual characters, and the contextual factors that influence this phenomenon. Considering the rising importance of social virtual reality, identifying and investigating these factors such as emotional affect help to enhance user experience, prevent the exploitation of human emotions, and enhance privacy in virtual spaces.

7. General Discussion

This thesis aims to explore the psychological factors that impact trust toward virtual humans within the context of virtual reality. The primary objectives of this thesis were twofold. Firstly, it aimed to identify a suitable behavioural tool for measuring interpersonal trust toward virtual humans in virtual reality. Secondly, it sought to utilize this tool, to examine the impact of psychological factors on a user's evaluation of a virtual human's trustworthiness under different psychological conditions.

7.1. Measuring Trust Toward Virtual Humans

To achieve the first research goal, the adjusted virtual reality version of the Social Conditioned Place Preference (SCPP: Kiser et al., 2022) paradigm was identified as a potential behavioural trust measurement tool in study 1 (see the introduction of Chapter 3 for a complete description of the SCPP paradigm). In study 1 (Chapter 3), an exploration of how the familiarity of a virtual human influences trust was assessed. Familiarity was of particular interest since interaction in social virtual reality is possible with familiar virtual humans (with a prior relationship or the appearance of the virtual human that resembles the user in real life) or unfamiliar virtual humans (with no prior relationship or the appearance is different to the user in real life) (Bombari et al., 2015; Latoschik et al., 2017). During the present study, participants were familiarized with one of two virtual humans, and they were tasked to explore a virtual art museum by looking at the art on the walls and approaching the virtual humans. The social place preference effect served as the main outcome measure, reflecting participants' preference for a specific room in the virtual art museum. Additionally, we examined variations in participants' evaluation of trust towards virtual humans and the resulting trust-related behaviours. These behaviours encompassed time spent in each room, proximity to virtual humans, information sharing with virtual humans, and response time to virtual humans inquiries. We expected participants to evaluate the familiar virtual human as more trustworthy, because they had more information to rely on when making a trust assessment, compared with the unfamiliar virtual human, of whom they had no prior information.

As previously stated, trust entails a degree of risk across various domains such as social, financial, personal, or organizational. According to the information processing theory, individuals make trust decisions by considering pertinent information (Wallace et al., 2003). This information can stem from previous experiences or, in the absence of such experiences, from personal tendencies (Wallace et al., 2003). An experimental study utilizing the basic decision-making task known as the Prisoner's Dilemma examined how participant's inclination to trust, predicts trustworthiness in both familiar and unfamiliar pairs over an extended period (Alarcon et al., 2018). The study found that propensity to trust was related to initial perceptions of trustworthiness in unfamiliar pairs but not in familiar pairs (Alarcon et al., 2018). Perceived trustworthiness between partners increased over time, indicating participants had an information-processing approach to assessing trustworthiness. Propensity to trust was related to initial assessments of trustworthiness but not to changes in perceptions over time. Familiarity with a partner influenced trust, with familiar pairs scoring higher on initial trustworthiness aspects, whereas unfamiliar pairs had lower perceived trustworthiness. Propensity to trust was only related to trust in unfamiliar pairs, while familiarity played a more important role in assessing trustworthiness (Alarcon et al., 2018).

Participants in our study evaluated both familiar and unfamiliar virtual humans as equally trustworthy. The repeated interaction with the virtual humans during the main task probably fostered a sense of familiarity with both. The sample size was also very small, making it challenging to observe behavioural effects in the proxies of trust. Nevertheless, the findings contradict traditional and virtual reality-based experimental studies that indicate a substantial influence of social interaction partners' familiarity on trust. Previous research suggests repeated interaction with an entity leads to familiarity, self-disclosure, higher levels of trust, and close proximity with the entity (Delton et al., 2011; Gulati & Sytch, 2008, Maloney et al., 2020; Rosenberger et al., 2020; Uzzi & Gillespie, 2002; Vugt et al., 2010). The insignificant results of the familiarity study mirror the results of Kiser et al. (2022). However, we question the suitability of the SCPP paradigm (Kiser et al., 2022) as a measure of trust behaviour toward virtual humans in virtual reality. The paradigm as a trust measurement tool needs validation before further use for this purpose and the proxies of trust need to be reassessed. The results on the familiarity of a virtual human on trust are deemed inconclusive due to the major limitations of the paradigm's validity. The results will not be discussed in depth. Nevertheless, the study was a meaningful and explorative attempt in the field of trust toward virtual humans in virtual reality. We continued to review the literature in search of an appropriate behavioural paradigm to measure trust toward a specific virtual human in virtual reality. This led to the construction of the virtual maze paradigm (Lin et al., 2023).

The virtual maze paradigm is inspired by the work of Hale et al. (2018) (for a comprehensive description of the paradigm, please refer to Chapter 2, specifically the section on measuring trust toward virtual humans). The construction of the virtual maze paradigm (Lin et al., 2023) had the academic virtual reality community in mind and considered fundamental virtual reality characteristics, such as presence, immersion, realism, and humanness (to name a few). Study 2 was conducted to verify whether our paradigm is sensitive to the differences in virtual human trustworthiness. Participants played one round of the established trust game (Berg et al., 1995) in virtual reality, they then completed the virtual maze task (see a full description of the paradigm in Chapter 4, study 2 or Lin et al. 2023). During the virtual maze task, participants supposedly had to escape from the maze by entering as few rooms as possible. After a supposed escape the task was completed, and a second round of the trust game was played as a control measure of the virtual maze. During the virtual maze, either a trustworthy or untrustworthy virtual human was present to provide navigation advice on request. The differences in the virtual human's trust evaluations and consequent trust behaviours (in the virtual maze: asking advice, following advice, time to respond to the given advice, and in the trust game: the number of tokens invested) were of interest. Significant differences between the advice asked and advice followed showed that the virtual maze is sensitive to differences in trust between virtual humans.

Furthermore, the significant differences in the second round of the trust game indicate that trust was built in the trustworthy condition and broken in the untrustworthy condition of the virtual maze task. The trust game can reflect trust-building during social interaction in the maze task. The virtual maze (Lin et al., 2023) fills the gap in the trust literature in virtual reality by contributing a flexible behavioural measurement tool suitable to measure trust towards virtual humans in virtual reality. Some general limitations of the paradigm include a lack of non-verbal social cues (e.g., body movements and vivid facial expressions corresponding to the vocals) that resulted in a medium level of virtual human social presence. Furthermore, the virtual maze was intentionally designed to allow simple decision-making processes and high levels of uncertainty by withholding feedback on progress in the maze. However, it could have led to participants losing motivation and mindless navigation through the virtual maze instead of elaborate decision-making. Due to the specific setup of the virtual maze and the cover story, this virtual maze 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 perceives a passing stranger's avatar as 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. During the design, fundamental virtual reality characteristics were considered, which make this virtual maze paradigm a novel contribution to the trust measurement tools.

7.2. Psychological Factors Influencing Trust

To achieve the second research goal, the virtual maze paradigm was utilized as an experimental task to investigate how psychological factors influence a user's evaluation of a virtual human's trustworthiness in various psychological conditions (Studies 3 and 4). In social virtual reality, users can do various tasks with other users, such as gaming, simulation or training, education, and learning, virtual tours or travel, social interactions, etc. (Roth et al., 2017). Some of these tasks can require a lot of cognitive resources, for example, educational tasks or doing a task for time (Bueno-Vesga et al., 2021; Trabucco et al., 2019). Users can experience different emotional states, that influence their behaviour and experience in the virtual world (Marín-Morales et al., 2020; Mesurado et al., 2021). In two experimental studies, the impact of cognitive load (high and low) (see Chapter 5, study 3 for the full description of the study design) and emotional affect (negative and positive) (see Chapter 6, study 4 for the full description of the study design) on trust was explored with the virtual maze task. In both studies, we measured participants' evaluation of the virtual human's trust and three behavioural proxies of trust: in the maze, 1) the number of times advice was asked, 2) the number of times advice was followed, and 3) the time it took participants to follow/execute the advice received. In study 3, the cognitive load of the virtual maze task was manipulated, and participants experienced both a high (including a secondary task) and a low (no added task) cognitive load condition. In study 4, the emotional affect of participants was manipulated with an autobiographical memory task, and participants experienced both a negative and a positive emotional affect condition. During the tasks in virtual reality participants were exposed to one picture and one sound (either positive or negative, depending on the condition) before entering a new room in the maze and before investing tokens in the trust game. During the maze task participants interacted with an allegedly human-controlled VHs (social interaction partners) during both experimental studies.

In Chapter 2 (see Figure 1), the trust process based on the work of Mayer et al. (1995) was discussed. We found that high cognitive load (in study 3) led to a higher trust evaluation of virtual humans. Users could use trust as a cognitive strategy to reduce cognitive complexity (Alarcon et al., 2018). To ensure the safety of participants during complex educational tasks or virtual environments with high cognitive complexity, designers need to consider these effects. The perception of time (Khawaji et al. 2014), repeated interaction (Cochard et al., 2004), and the perceived helpfulness of the virtual human (Klümper & Sürth, 2023) could also contribute towards trust in a virtual human during high cognitive load or emergency situations. Appropriate time must be allowed during complex tasks to enable users to process information critically to prevent impulsive decision-making. Repeated interaction with a virtual human under cognitive constraints increases the risk of users evaluating an untrustworthy virtual human as a reliable source of social support without accurately assessing the trustworthiness or considering potential consequences (Brown, 1997; Cochard et al., 2004; Klümper & Sürth, 2023; Kothgassner et al., 2019). By avoiding impulsive acceptance of untrustworthy virtual humans, users can protect themselves, make informed decisions, uphold trustworthiness standards, and preserve their autonomy within virtual environments (Adams et al., 2018; Burnett Heyes et al., 2012; Lake, 2020; Lin & Latoschik, 2022). Participants in study 4 interacted with a virtual human who was trustworthy in appearance (a medical doctor), which explains the high trust evaluation. Participant's behaviour in study 4 was probably due to their trust in the virtual human and not because of the emotional affect. However, evaluating the virtual human's trustworthiness does not necessarily mean a person would also behave in a way that indicates they trust the virtual human.

As summarized in the theoretical background (Chapter 2) subjective ratings are poor predictors of external behaviour (Armitage & Christian, 2003; Glaeser et al., 2000; McCambridge et al., 2012). Studies 3 and 4 demonstrate that despite the differences in the virtual human's evaluation of trust between conditions, participants asked and followed advice equally often. It is possible that asking for advice could be perceived as the most obvious choice.

The order in which the emotional affect conditions were presented to the participants in Study 4 influenced their behaviour significantly. When the negative emotional affect condition was presented as the first condition, fewer trust behaviour were observed. It could be that the initial evaluation of the virtual human as a stranger negatively influenced the perceived helpfulness of the virtual human (Sandstrom et al., 2022). It is possible that when users do a task, enter a virtual environment, or interact with an unfamiliar virtual human for the first time higher levels of cognitive load and negative emotional affect can be experienced, compared with when they do so for the second time. The increase of uncertainty and negative emotional affect accompanied by new experiences negatively influence the availability of cognitive resources and result in users behaving impulsively without considering the consequences of their behaviour (Eben et al., 2020; Harmon-Jones et al., 2013; Plass et al., 2010; Sadeh & Bredemeier, 2021). Considering how cognitive load and emotional affect influence user's behaviour, when designing social virtual reality applications can enhance the user performance and experience, minimize cognitive overload and reduce impulsivity. For example, if the impact of cognitive load on trust is considered in the design of educational applications the comprehension, retention, and transfer of knowledge or skills can be enhanced, while learners have a more enjoyable learning experience. Managing the effects of cognitive load on impulsive decision-making is paramount for user’s safety and performance in real-world applications, especially in high-stress occupations or critical decision-making tasks. Understanding and mitigating the effect of cognitive load in critical situations or complex tasks in virtual reality can minimize potential risks, errors, or performance decrements, ensuring better outcomes in professional and real-world applications. Designing social virtual reality applications with the consideration of how emotional affect influences users' behaviour is crucial. For example, when using social virtual reality as a mental health tool, emotional affect significantly impacts empathy and the sense of connection between individuals (Cohen et al., 2021; Deighan et al., 2023; Van Kleef et al., 2016).

Therefore, the results of Study 4 (in Chapter 6) need to be considered when designing social virtual reality applications that evoke emotional responses in users. The study's findings can inform the design of virtual humans in social virtual reality environments that need to convey trustworthiness, i.e., in virtual reality healthcare applications. Developers can consider the impact of negative emotional affect on users’ trust and reaction times, aiming to create virtual humans and experiences that elicit positive emotions and foster trust. The study's findings highlight the importance of considering the consequence of emotional experiences in virtual reality interactions. If negative emotional affects are encountered early on, they may have a lasting effect on trust and subsequent behaviours. Designers can explore strategies for introducing positive emotional states at the onset of virtual interactions to establish a foundation of trust. The findings can inform training programs and interventions involving virtual humans in social virtual reality. For example, if the goal is to enhance trust and reaction times in certain professional settings, such as virtual job interviews or therapy sessions, it may be beneficial to incorporate techniques that promote positive emotional states before introducing potentially stressful or negative situations. Furthermore, appropriate emotional responses to users’ current emotional state by computer-controlled virtual humans can enhance user's ability to empathize with virtual characters, leading to deeper and more meaningful social interactions (for example when using social virtual reality to encourage prosocial behaviour (Rosenberg et al., 2013)). Furthermore, considering human vulnerabilities such as user trust, impulsivity, emotions, lack of cognitive resources, and uncertainty in virtual reality applications prevent their exploitation and protect users against cybercriminals (Kadena & Gupi, 2021) (see Table 12, for a summary of the key findings of the experimental studies).

Table 12: Brief Summary of Study FindingsCognitve Load:

Chapter	Study	Research Goals	N	Experimental Design	Main findings
3	1: Who you meet: *Familiarity* Study	Testing a potential behavioural paradigm to measure interpersonal trust in VR & investigate the impact of familiarity of a VH on interpersonal trust	20	Within-subject design	The non-significant results cast doubt on the suitability of the SCPP paradigm as a measure of trust behaviour. The manipulation of familiarity may not have been strong enough to influence participants' perceptions of trustworthiness in the VHs and consequent trust behaviour. Thus, the results of familiarity on trust are *inconclusive*.
4	2: The Virtual Maze: Validat*ion* Study	Construction and validation of a behavioural paradigm to measure interpersonal trust toward VHs in VR	70	Between-subject design	The Virtual Maze paradigm is *sensitive* to the *manipulation of trustworthiness* in VH's. We contribute a *novel behavioural paradigm* that measures interpersonal trust toward VHs in VR.
5	3: What you do: *Cognitive Load* Study	Investigating how cognitive load (high and low) impact interpersonal trust toward a VH in VR with the Virtual Maze paradigm.	50	Within-subject design	CL did not have a significant impact on trust toward the VH. Due to a lack of available *cognitive resources* in the HCL condition, participants could not critically process information and *responded to advice* more impulsively (significantly *faster*). *Order effect*: when LCL was presented as the first condition participants r*esponded to advice significantly faster,* compared to the HCL presented as the first condition.
6	4: How we feel: *Emotional Affect* Study	Investigating how emotional affect (negative and positive) impact interpersonal trust toward a VH in VR with the Virtual Maze paradigm.	50	Within-subject design	EA did not significantly impact trust toward the VH; the VH was evaluated as equally trustworthy in both conditions. No significant differences in advice asked or followed between conditions were observed. Participants *responded to advice faster* in the NEA condition. *Order effect*: when NEA was presented as the first condition participants *asked and followed advice significantly less*.

Note: SCPP; Social Conditioned Place Preference, VR; Virtual Reality, VH; Virtual Human, CL; Cognitive Load, HCL; High Cognitive Load, LCL; Low Cognitive Load, EA; Emotional Affect, NEA; Negative Emotional Affect.

7.3. Limitations and Outlook

Interpersonal trust is essential for building and maintaining relationships in the real and virtual worlds (Caldwell & Clapham, 2003). The acceptance of social media communication and virtual reality led to significant investments in the "Metaverses" by tech giants such as *Meta*, *Nvidia*, and *Microsoft* (Kim, 2021). In line with this trend, there is good reason to be convinced that the multiuser virtual reality application of social virtual reality has great potential to lead the next revolution in the digitalization of social activities (O'Brolcháin et al., 2016). Amidst the exhilaration and progress surrounding virtual reality technology, there is a growing apprehension that the rapid development lacks consideration of human risk factors (Easa, 2021; O'Brolcháin et al., 2016). While the virtual realm offers unprecedented possibilities and immersive experiences, the potential consequences of the misuse of human factors should not be underestimated or overlooked (Easa, 2021; O'Brolcháin et al., 2016). Neglecting to address these inherent risks endangers users and undermines the long-term viability and acceptance of this transformative technology. To ensure a responsible future for this transformative technology, we must prioritize comprehensive assessments of human factors and safety considerations. Users of social virtual reality's privacy are vulnerable in the virtual world, as introduced in Chapter 1. In virtual reality, user's trust can be exploited through immersive experiences that seem real, misrepresentation by malicious actors, and deception can lead to harmful activities (Jansen & Fischbach, 2020). Just as in the real world, individuals can be manipulated through social engineering techniques in virtual reality (Mouton et al., 2016). Scammers or malicious actors can attempt to gain the trust of users by posing as someone else, tricking users into revealing personal information, or persuading them to take actions that could be harmful. Exploiting a user's trust could further involve intentionally targeting vulnerable individuals or using deceptive tactics to manipulate others' emotions or behaviours (Mouton et al., 2016).

The experimental studies have limitations. A general limitation of studies 3 and 4 is that only a trustworthy virtual human and no untrustworthy virtual human were included. A between-subject design with one trustworthy and one untrustworthy virtual human would be interesting to compare results between interaction with a trustworthy and an untrustworthy virtual human. Furthermore, due to the absence of a standardized subjective measurement tool that measures trust towards virtual humans in virtual reality, we relied on self-constructed questions inspired by the literature that have yet to be validated. Future studies could focus on the creation and standardization of a virtual reality-specific virtual human trust scale that considers an individual's literacy of virtual reality and virtual humans, as well as incorporating virtual reality-specific characteristics. Avatar literacy refers to an individual's knowledge, skill, and understanding regarding the creation, customization, and effective use of avatars in virtual environments (Yildiz et al., 2019). Avatar-literate individuals understand the nuances of avatar design and behaviour and can evaluate the authenticity and credibility of virtual humans (Kolesnichenko et al., 2019; Yu et al., 2016). Furthermore, being aware of techniques used to create or customize false avatar appearances or behaviours allows individuals to accurately assess whether the virtual human is genuinely trustworthy or attempting to deceive them. Avatar literacy is thus a skill necessary for accurate trust assessments to ensure individuals navigate and participate effectively in virtual environments, leveraging the full potential of avatars for communication, self-expression, and meaningful interactions within virtual communities.

In the virtual maze, the colours of the doors were randomized, presenting participants with doors of either the same or different colours. Another general limitation of the experimental studies was that the colours of the doors were not assessed in a pre-study before the virtual maze was designed. This limitation could have influenced participants' advice asking, and following behaviour during the studies as this was not investigated when the virtual maze was validated (Kaur, 2020; Silic & Cyr, 2016). For example, an experimental study investigating the relationship between colours and emotions found that humans associate colours with emotions (Kaur, 2020). This observation indicated a positive emotional association with bright colours and a negative emotional association with dark colours (Kaur, 2020). Participants' preferences could have been influenced rather by the colour of the door than their propensity to trust the virtual human. Alternatively, the colours of the doors could have influenced participants; this effect was not measured. There were other study-specific limitations that we addressed in the respective chapters such as a lack of a pre-study to select a virtual human and proxies of trust in the familiarity study. In the virtual maze validation study (Study 2) the social presence of the virtual humans was at a medium level, possibly due to a lack of non-verbal cues, a lack of feedback, and potential loss of motivation in the virtual maze. Only female participants were recruited for the experimental studies in this thesis; it would be interesting to replicate the studies with male participants to explore gender differences in interpersonal trust. Traditional experiments found gender differences in trust-related behaviours and preferences, suggesting that men and women perceive and exhibit trust differently (Mattarozzi et al., 2015; Rau, 2012; Wu et al., 2020). It would be interesting to explore whether this gender difference exists in virtual reality with male and female virtual humans. Finally, in study 4 only a trustworthy virtual human was included in the study. Future research can improve the virtual maze by removing or controlling potential confounding factors such as the colours of the doors or including two virtual humans (trustworthy vs. untrustworthy) in within-subject experimental designs that investigate how psychological factors influence trust. The limitations were discussed in detail in the respective studies.

7.4. Conclusion

This thesis examines the impact of psychological factors on trust toward virtual humans in virtual reality. Key contributions include the impact of virtual human familiarity on trust evaluation and behaviour, the development of a behavioural trust measurement tool (the virtual maze paradigm), and the use of this tool to investigate the impact of cognitive load and emotional affect on trust toward virtual humans in virtual reality. Consistent with other experimental studies, the studies in this thesis found that subjective evaluations of virtual human trustworthiness may not always align with corresponding trust behaviour (Armitage & Christian, 2003; McCambridge et al., 2012). The familiarity study reveals that participants evaluated both familiar and unfamiliar virtual humans as equally trustworthy and no significant behavioural differences were observed. Due to the non-significant results, limitations of the paradigm, and lack of validation of the proposed proxies of trust the results of this study are deemed inconclusive. In study 2, the virtual maze paradigm demonstrated its sensitivity to variations in virtual human trustworthiness. This was reflected in significant differences observed in both subjective evaluations and behavioural responses among participants. The virtual maze paradigm was used to investigate the effects of cognitive load and emotional affect on virtual human trust evaluations in two experimental studies. The manipulation of cognitive load and emotional affect did not impact the subjective evaluation of the virtual human's trustworthiness. In study 3, participants responded to advice significantly faster during the high cognitive load condition. Furthermore, when the low cognitive load condition was presented first, participants experienced more cognitive load and, as a result, responded to advice significantly faster. In study 4 participants responded to advice faster while experiencing a negative emotional affect. Furthermore, when the negative emotional affect condition was experienced first, participants asked and followed advice more. These findings emphasize the importance of considering cognitive and emotional factors in the trust assessment of virtual humans. Overall, this thesis offers a novel tool for trust measurement (the virtual maze paradigm), and contributes to understanding psychological factors in virtual human trust in virtual reality.

References

Abatayo, A. L., Lynham, J., & Sherstyuk, K. (2020). Communication, Expectations, and Trust: An Experiment with Three Media. *Games*, *11*(4), 48. https://doi.org/10.3390/g11040048

Achenbach, J., Waltemate, T., Latoschik, M. E., & Botsch, M. (2017). Fast generation of realistic virtual humans. *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, 1–10. https://doi.org/10.1145/3139131.3139154

Adams, D., Bah, A., & Barwulor, C. (2018). *Ethics Emerging: The Story of Privacy and Security Perceptions in Virtual Reality*. 17.

Adcock, A. (2000). *Effects of Cognitive Load on Processing and Performance*.

Agarwal, M., Mehra, M., Pawar, R., & Shah, D. (2011). Secure authentication using dynamic virtual keyboard layout. *Proceedings of the International Conference & Workshop on Emerging Trends in Technology - ICWET '11*, 288. https://doi.org/10.1145/1980022.1980087

Ahmad, M. I., Bernotat, J., Lohan, K., & Eyssel, F. (2019). *Trust and Cognitive Load During Human-Robot Interaction* (arXiv:1909.05160). arXiv. https://doi.org/10.48550/arXiv.1909.05160

Alarcon, G. M., Lyons, J. B., Christensen, J. C., Klosterman, S. L., Bowers, M. A., Ryan, T. J., Jessup, S. A., & Wynne, K. T. (2018). The effect of propensity to trust and perceptions of trustworthiness on trust behaviors in dyads. *Behavior Research Methods*, *50*(5), 1906–1920. https://doi.org/10.3758/s13428-017-0959-6

Allport, A. (1985). The historical background of social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 1, pp. 1–46). Random House.

Alós-Ferrer, C., & Farolfi, F. (2019). Trust Games and Beyond. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00887

Alsulaiman, F., & Saddik, A. (2006). A Novel 3D Graphical Password Schema. *2006 IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 125–128. https://doi.org/10.1109/VECIMS.2006.250805

Andreatta, M., Neueder, D., Herzog, K., Genheimer, H., Schiele, M. A., Deckert, J., Domschke, K., Reif, A., Wieser, M. J., & Pauli, P. (2020). Generalization of

Conditioned Contextual Anxiety and the Modulatory Effects of Anxiety Sensitivity. *Neurotherapeutics*, *17*(3), 1239–1252. https://doi.org/10.1007/s13311-020-00831-8

Andreatta, M., & Pauli, P. (2015). Appetitive vs. Aversive conditioning in humans. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00128

Argelaguet, F., Hoyet, L., Trico, M., & Lecuyer, A. (2016). The role of interaction in virtual embodiment: Effects of the virtual hand representation. *2016 IEEE Virtual Reality (VR)*, 3–10. https://doi.org/10.1109/VR.2016.7504682

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

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

Atlas, S., & Putterman, L. (2011). Trust among the Avatars: A Virtual World Experiment, with and without Textual and Visual Cues. *Southern Economic Journal*, *78*(1), 63–86. https://doi.org/10.4284/0038-4038-78.1.63

Babel, F., Kraus, J., Miller, L., Kraus, M., Wagner, N., Minker, W., & Baumann, M. (2021). Small Talk with a Robot? The Impact of Dialog Content, Talk Initiative, and Gaze Behavior of a Social Robot on Trust, Acceptance, and Proximity. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-020-00730-0

Bagneux, V., Bollon, T., & Dantzer, C. (2012). Do (un)certainty appraisal tendencies reverse the influence of emotions on risk taking in sequential tasks? *Cognition & Emotion*, *26*(3), 568–576. https://doi.org/10.1080/02699931.2011.602237

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

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin*, *29*(7), 819–833. https://doi.org/10.1177/0146167203029007002

Baker, C., Pawling, R., & Fairclough, S. (2020). Assessment of threat and negativity bias in virtual reality. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-74421-1

Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939

Baron, D., Holland, C. M., Carlson, K., Wolfrum, E., & Thompson, B. L. (2020). Adapting social conditioned place preference for use in young children. *Neurobiology of Learning and Memory*, *172*, 107235. https://doi.org/10.1016/j.nlm.2020.107235

Becker, L. C. (1996). Trust as Noncognitive Security about Motives. *Ethics*, *107*(1), 43–61. https://doi.org/10.1086/233696

Bee, K., Bee, N., Andre, E., Berkmüller, U., & Wagner, J. (2011). Physiological Measurement of Trust-Related Behavior in Trust-Neutral and Trust-Critical Situations. In *IFIP Advances in Information and Communication Technology* (Vol. 358, p. 172). https://doi.org/10.1007/978-3-642-22200-9_14

Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2012). *Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen (KUSIV3)*.

Bejarano, H., Gillet, J., & Lara, I. R. (2021). *When the rich do (not) trust the (newly) rich: Experimental evidence on the effects of positive random shocks in the trust game* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/wmejt

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). *Who is "Behavioral"? Cognitive ability and anomalous preferences*. *11*, 1231–1255.

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79.

Bente, G., Dratsch, T., Rehbach, S., Reyl, M., & Lushaj, B. (2014). Do You Trust My Avatar? Effects of Photo-Realistic Seller Avatars and Reputation Scores on Trust in Online Transactions. In F.F.-H. Nah (Ed.), *HCI in Business* (pp. 461–470). Springer International Publishing.

Bente, G., Rüggenberg, S., & Krämer, N. (2004). *Social presence and interpersonal trust in avatar-based, collaborative net-communications*.

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

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Biegelman, M. T. (2013). *Faces of fraud: Cases and lessons from a life of fighting fraud*. John Wiley & Sons, Inc.

Biros, D. P., Daly, M., & Gunsch, G. (2004). The Influence of Task Load and Automation Trust on Deception Detection. *Group Decision and Negotiation*, *13*(2), 173–189. https://doi.org/10.1023/B:GRUP.0000021840.85686.57

Blackwell, L., Ellison, N., Elliott-Deflo, N., & Schwartz, R. (2019). Harassment in Social Virtual Reality: Challenges for Platform Governance. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 100:1-100:25. https://doi.org/10.1145/3359202

Blascovich, J. (2002). A theoretical model of social influence for increasing the utility of collaborative virtual environments. *Proceedings of the 4th International Conference on Collaborative Virtual Environments*, 25–30. https://doi.org/10.1145/571878.571883

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). TARGET ARTICLE: Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Blender Foundation. (n.d.). *Blender.org—Home of the Blender project—Free and Open 3D Creation Software*. Blender.Org. Retrieved October 9, 2022, from https://www.blender.org/

Block, R. A., George, E. J., & Reed, M. A. (1980). A watched pot sometimes boils: A study of duration experience. *Acta Psychologica*, *46*(2), 81–94. https://doi.org/10.1016/0001-6918(80)90001-3

Block, R. A., & Gruber, R. P. (2014). Time perception, attention, and memory: A selective review. *Acta Psychologica*, *149*, 129–133. https://doi.org/10.1016/j.actpsy.2013.11.003

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*(3), 330–343. https://doi.org/10.1016/j.actpsy.2010.03.006

Bohner, G., Ruder, M., & Erb, H.-P. (2002). When expertise backfires: Contrast and assimilation effects in persuasion. *The British Journal of Social Psychology*, *41*(Pt 4), 495–519. https://doi.org/10.1348/014466602321149858

Bolton, G. E., Katok, E., & Ockenfels, A. (2004). Trust among Internet Traders. *Analyse & Kritik*, *26*(1), 185–202. https://doi.org/10.1515/auk-2004-0110

Bombari, D., Schmid Mast, M., Canadas, E., & Bachmann, M. (2015). Studying social interactions through immersive virtual environment technology: Virtues, pitfalls, and future challenges. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00869

Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General*, *142*(1), 143–150. https://doi.org/10.1037/a0028930

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276–298. https://doi.org/10.1037/1528-3542.1.3.276

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Breuer, C., Hüffmeier, J., Hibben, F., & Hertel, G. (2020). Trust in teams: A taxonomy of perceived trustworthiness factors and risk-taking behaviors in face-to-face and virtual teams. *Human Relations*, *73*(1), 3–34. https://doi.org/10.1177/0018726718818721

Breyer, B., & Bluemke, M. (2016). Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/ZIS242

Breyer,B. (2015). *Compilation of social science items and scales (ZIS)*. https://doi.org/10.6102/zis235

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, *52*(2), 455–463. https://doi.org/10.3758/s13428-019-01242-0

Brown, H. G., Poole, M. S., & Rodgers, T. L. (2004). Interpersonal Traits, Complementarity, and Trust in Virtual Collaboration. *Journal of Management Information Systems*, *20*(4), 115–138. https://doi.org/10.1080/07421222.2004.11045785

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*(7), 1118–1140. https://doi.org/10.3758/bf03205526

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 53–61. https://doi.org/10.1207/S15326985EP3801_7

Bueno-Vesga, J. A., Xu, X., & He, H. (2021). The Effects of Cognitive Load on Engagement in a Virtual Reality Learning Environment. *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 645–652. https://doi.org/10.1109/VR50410.2021.00090

Burnett Heyes, S., Adam, R. J., Urner, M., Van Der Leer, L., Bahrami, B., Bays, P. M., & Husain, M. (2012). Impulsivity and Rapid Decision-Making for Reward. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00153

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, *76*, 839–855. https://doi.org/10.1037/0022-3514.76.5.839

Calcagnetti, D. J., & Schechter, M. D. (1992). Place conditioning reveals the rewarding aspect of social interaction in juvenile rats. *Physiology & Behavior*, *51*(4), 667–672. https://doi.org/10.1016/0031-9384(92)90101-7

Caldwell, C., & Clapham, S. E. (2003). Organizational Trustworthiness: An International Perspective. *Journal of Business Ethics*, *47*(4), 349–364. https://doi.org/10.1023/A:1027370104302

Camerer, C., F. (2010). Behavioural game theory. In S. N. Durlauf & L. E. Blume (Eds.), *Behavioural and Experimental Economics* (pp. 42–50). Palgrave Macmillan UK. https://doi.org/10.1057/9780230280786_6

Camerer, C., & Weigelt, K. (1988). Experimental Tests of a Sequential Equilibrium Reputation Model. *Econometrica*, *56*(1), 1. https://doi.org/10.2307/1911840

Caprara, M., Gerbino, M., Mebane, M. E., & Ramirez-Uclés, I. M. (2022). Self-efficacy beliefs in managing positive emotions: Associations with positive affect, negative affect, and life satisfaction across gender and ages. *Frontiers in Human Neuroscience*, *16*, 927648. https://doi.org/10.3389/fnhum.2022.927648

Cardoso, R. C., Gomes, A. J. P., & Freire, M. M. (2017). A User Trust System for Online Games-Part I: An Activity Theory Approach for Trust Representation. *Ieee Transactions on Computational Intelligence and Ai in Games*, *9*(3), 305–320. https://doi.org/10.1109/TCIAIG.2016.2592965

Chan, D. (2008). So Why Ask Me? Are Self-Report Data Really That Bad? In *Statistical and Methodological Myths and Urban Legends*. Routledge.

Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*(2), 87–105. https://doi.org/10.1016/j.cogpsych.2010.03.001

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust Multimodal Cognitive Load Measurement*. Springer International Publishing. https://doi.org/10.1007/978-3-319-31700-7

Chiou, E. K., Schroeder, N. L., & Craig, S. D. (2020). How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education*, *146*, 103756. https://doi.org/10.1016/j.compedu.2019.103756

Chun, K.-T., & Campbell, J. B. (1974). Dimensionality of the Rotter Interpersonal Trust Scale. *Psychological Reports*, *35*(3), 1059–1070. https://doi.org/10.2466/pr0.1974.35.3.1059

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

Cochard, F., Nguyen Van, P., & Willinger, M. (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior & Organization*, *55*(1), 31–44. https://doi.org/10.1016/j.jebo.2003.07.004

Cohen, D., Landau, D. H., Friedman, D., Hasler, B. S., Levit-Binnun, N., & Golland, Y. (2021). Exposure to social suffering in virtual reality boosts compassion and facial synchrony. *Computers in Human Behavior*, *122*, 106781. https://doi.org/10.1016/j.chb.2021.106781

Couch, L. L., & Jones, W. H. (1997). Measuring Levels of Trust. *Journal of Research in Personality*, *31*(3), 319–336. https://doi.org/10.1006/jrpe.1997.2186

Dai, B., Zhang, W., Wang, Y., & Jian, X. (2020). Comparison of Trust Assessment Scales Based on Item Response Theory. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00010

Danaher, J., & Sætra, H. S. (2022). Technology and moral change: The transformation of truth and trust. *Ethics and Information Technology*, *24*(3), 35. https://doi.org/10.1007/s10676-022-09661-y

Davidson, R. J., Fox, A., & Kalin, N. H. (2007). *Handbook of emotion regulation*.

De Borst, A. W., & De Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00576

De La Rosa, S., & Breidt, M. (2018). Virtual reality: A new track in psychological research. *British Journal of Psychology*, *109*(3), 427–430. https://doi.org/10.1111/bjop.12302

de Visser, E. J. (2012). *THE WORLD IS NOT ENOUGH: TRUST IN COGNITIVE AGENTS*.

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

De Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': The importance of trust repair in human–machine interaction. *Ergonomics*, *61*(10), 1409–1427. https://doi.org/10.1080/00140139.2018.1457725

Deighan, M. T., Ayobi, A., & O'Kane, A. A. (2023). Social Virtual Reality as a Mental Health Tool: How People Use VRChat to Support Social Connectedness and Wellbeing. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3544548.3581103

del Angel, J., Cortez, J., Juárez, D., Guerrero, M., García, A., Ramírez, C., & Valdez, P. (2015). Effects of sleep reduction on the phonological and visuospatial components of working memory. *Sleep Science*, *8*(2), 68–74. https://doi.org/10.1016/j.slsci.2015.06.001

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, *108*(32), 13335–13340. https://doi.org/10.1073/pnas.1102131108

Derlega, V. J., & Chaikin, A. L. (1977). Privacy and Self-Disclosure in Social Relationships. *Journal of Social Issues*, *33*(3), 102–115. https://doi.org/10.1111/j.1540-4560.1977.tb01885.x

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

Duffy, S., & Smith, J. (2012, May 15). *Cognitive load in the multi-player prisoner's dilemma game: Are there brains in games?* [MPRA Paper]. https://mpra.ub.uni-muenchen.de/53123/

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and Believing: The Influence of Emotion on Trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, C., Jebabli, I., … Wamba, S. F. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *66*, 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

Eagly, A. H., & Wood, W. (2012). Handbook of Theories of Social Psychology: Volume 2. In *Handbook of Theories of Social Psychology: Volume 2* (1–2, pp. 458–476). SAGE Publications Ltd. https://doi.org/10.4135/9781446249222

Easa, S. (2021). Human Factor Considerations in Virtual Reality: Adequate or Inadequate? *Ergonomics International Journal*, *5*(2). https://doi.org/10.23880/eoij-16000267

Eben, C., Billieux, J., & Verbruggen, F. (2020). Clarifying the Role of Negative Emotions in the Origin and Control of Impulsive Actions. *Psychologica Belgica*, *60*(1), 1–17. https://doi.org/10.5334/pb.502

Echterhoff, G., Higgins, E. T., & Groll, S. (2005). Audience-tuning effects on memory: The role of shared reality. *Journal of Personality and Social Psychology*, *89*(3), 257–276. https://doi.org/10.1037/0022-3514.89.3.257

Eggemeier, F. T., & Wilson, G. F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In *Multiple-task performance*. CRC Press.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519. https://doi.org/10.1162/jocn.2007.19.9.1508

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*(6), 1585–1593. https://doi.org/10.1016/j.jrp.2008.07.011

Falk, B., Meng, Y., Zhan, Y., & Zhu, H. (2021). POSTER: ReAvatar: Virtual Reality De-anonymization Attack Through Correlating Movement Signatures. *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2405–2407. https://doi.org/10.1145/3460120.3485345

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fehr, E. (2009). On The Economics and Biology of Trust. *Journal of the European Economic Association*, *7*(2–3), 235–266. https://doi.org/10.1162/JEEA.2009.7.2-3.235

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A.-K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human-Computer Studies*, *82*, 48–56. https://doi.org/10.1016/j.ijhcs.2015.05.004

Fiedler, M., & Haruvy, E. (2009). The lab versus the virtual lab and virtual field—An experimental investigation of trust games with communication. *Journal of Economic Behavior & Organization*, *72*(2), 716–724. https://doi.org/10.1016/j.jebo.2009.07.013

Freeman, G., & Maloney, D. (2021). 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. https://doi.org/10.1145/3432938

Freeman, G., Zamanifard, S., Maloney, D., & Adkins, A. (2020). My Body, My Avatar: How People Perceive Their Avatars in Social Virtual Reality. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8. https://doi.org/10.1145/3334480.3382923

George, C., Eiband, M., Hufnagel, M., & Hussmann, H. (2018). Trusting Strangers in Immersive Virtual Reality. *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, 1–2. https://doi.org/10.1145/3180308.3180355

Gigante, M. A. (1993). Virtual Reality: Definitions, History and Applications. In *Virtual Reality Systems* (pp. 3–14). Elsevier. https://doi.org/10.1016/B978-0-12-227748-1.50009-3

Gilbert, D., T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. Uleman S. & J. Bargh A. (Eds.), *Unintended thoughts* (pp. 189–211). Guilford Press. https://acmelab.yale.edu/sites/default/files/1999_the_unbearable_automaticity_of_being.pdf

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*(5), 733–740. https://doi.org/10.1037/0022-3514.54.5.733

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring Trust*. *The Quarterly Journal of Economics*, *115*(3), 811–846. https://doi.org/10.1162/003355300554926

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Golden, S. A., Aleyasin, H., Heins, R., Flanigan, M., Heshmati, M., Takahashi, A., Russo, S. J., & Shaham, Y. (2017). Persistent conditioned place preference to aggression experience in adult male sexually-experienced CD-1 mice: Persistent aggression conditioned place preference in CD-1 mice. *Genes, Brain and Behavior*, *16*(1), 44–55. https://doi.org/10.1111/gbb.12310

Gombolay, M., Yang, X. J., Hayes, B., Seo, N., Liu, Z., Wadhwania, S., Yu, T., Shah, N., Golen, T., & Shah, J. (2018). Robotic assistance in the coordination of patient care. *The International Journal of Robotics Research*, *37*(10), 1300–1316. https://doi.org/10.1177/0278364918778344

Gonzalez-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., Maselli, A., Banakou, D., Pelechano, N., Orts-Escolano, S., Orvalho, V., Trutoiu, L., Wojcik, M., Sanchez-Vives, M. V., Bailenson, J., Slater, M., & Lanier, J. (2020). The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Frontiers in Virtual Reality*, *1*. https://www.frontiersin.org/articles/10.3389/frvir.2020.561558

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

Greenlees, I., Buscombe, R., Thelwell, R., Holder, T., & Rimmer, M. (2005). Impact of Opponents' Clothing and Body Language on Impression Formation and Outcome Expectations. *Journal of Sport and Exercise Psychology*, *27*(1), 39–52. https://doi.org/10.1123/jsep.27.1.39

Gromer, D., Madeira, O., Gast, P., Nehfischer, M., Jost, M., Müller, M., Mühlberger, A., & Pauli, P. (2018). Height Simulation in a Virtual Reality CAVE System: Validity of Fear Responses and Effects of an Immersion Manipulation. *Frontiers in Human Neuroscience*, *12*, 372. https://doi.org/10.3389/fnhum.2018.00372

Gulati, R., & Sytch, M. (2008). Does Familiarity Breed Trust? Revisiting the Antecedents of Trust. *Managerial and Decision Economics*, *29*(2/3), 165–190.

Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., & Billinghurst, M. (2019). In AI We Trust: Investigating the Relationship between Biosignals, Trust and Cognitive Load in VR. *25th ACM Symposium on Virtual Reality Software and Technology*, 1–10. https://doi.org/10.1145/3359996.3364276

Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., & Billinghurst, M. (2020). Measuring Human Trust in a Virtual Assistant using Physiological Sensing in Virtual

Reality. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 756–765. https://doi.org/10.1109/VR46266.2020.00099

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., & Puranen, B. (2020). *World Values Survey Wave 7 (2017-2020) Cross-National Data-Set* (1.5) [dataset]. World Values Survey Association. https://doi.org/10.14281/18241.1

Hale, J., & Hamilton, A. F. D. C. (2016). Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep35295

Hale, J., Payne, M. E., Taylor, K. M., Paoletti, D., & De C Hamilton, A. F. (2018). The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology*, *71*(4), 989–1008. https://doi.org/10.1080/17470218.2017.1307865

Hall, E. T., & Hall, E. T. (1966). *The hidden dimension* (Vol. 609). Anchor.

Han, T. A., Perret, C., & Powers, S. T. (2021). An evolutionary game theory analysis of trust in repeated games and lessons for human-machine interactions. *The 2021 Conference on Artificial Life*. The 2021 Conference on Artificial Life, Online. https://doi.org/10.1162/isal_a_00417

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Harmon-Jones, E., Gable, P. A., & Price, T. F. (2013). Does Negative Affect Always Narrow and Positive Affect Always Broaden the Mind? Considering the Influence of Motivational Intensity on Cognitive Scope. *Current Directions in Psychological Science*, *22*(4), 301–307. https://doi.org/10.1177/0963721413481353

Hernandez, I., & Preston, J. L. (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology*, *49*(1), 178–182. https://doi.org/10.1016/j.jesp.2012.08.010

Hertzum, M., & Holmegaard, K. D. (2013). Perceived Time as a Measure of Mental Workload: Effects of Time Constraints and Task Success. *International Journal of Human-Computer Interaction*, *29*(1), 26–39. https://doi.org/10.1080/10447318.2012.676538

Higgins, D., Zibrek, K., Cabral, J., Egan, D., & McDonnell, R. (2022). Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. *Computers & Graphics, 104*, 116–128. https://doi.org/10.1016/j.cag.2022.03.009

Hiller, L. T., Takata, S., & Thompson, B. L. (2015). Conditioned place preference successfully established in typically developing children. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00187

Hinson, J. M., Jameson, T. L., & Whitney, P. (2002). Somatic markers, working memory, and decision making. *Cognitive, Affective & Behavioral Neuroscience*, *2*(4), 341–353. https://doi.org/10.3758/cabn.2.4.341

Iacono, C. S., & Weisband, S. (1997). Developing trust in virtual teams. *Proceedings of the Thirtieth Hawaii International Conference on System Sciences*, *2*, 412–420. https://doi.org/10.1109/HICSS.1997.665615

IADS-E. (n.d.). Retrieved May 23, 2023, from https://sites.google.com/view/iads-e/

International Affective Picture System (IAPs) | Research Administration. (n.d.). Retrieved May 23, 2023, from https://www.umass.edu/research/guidance/international-affective-picture-system-iaps

Jamieson, L. (2007). Intimacy. In G. Ritzer (Ed.), *The Blackwell Encyclopedia of Sociology* (p. wbeosi071). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405165518.wbeosi071

Jansen, P., & Fischbach, F. (2020). The Social Engineer: An Immersive Virtual Reality Educational Game to Raise Social Engineering Awareness. *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*, 59–63. https://doi.org/10.1145/3383668.3419917

Jasielska, D. (2020). The moderating role of kindness on the relation between trust and happiness. *Current Psychology*, *39*(6), 2065–2073. https://doi.org/10.1007/s12144-018-9886-7

Jasielska, D., Rogoza, R., Zajenkowska, A., & Russa, M. B. (2019). General trust scale: Validation in cross-cultural settings. *Current Psychology*. https://doi.org/10.1007/s12144-019-00435-2

Johnson, K., Lennon, S. J., & Rudd, N. (2014). Dress, body and self: Research in the social psychology of dress. *Fashion and Textiles, 1*(1), 20. https://doi.org/10.1186/s40691-014-0020-7

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Johnson-George, C., & Swap, W. C. (1982). 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*(6), 1306–1317. http://dx.doi.org/10.1037/0022-3514.43.6.1306

Johnson-Glenberg, M. C. (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. *Frontiers in Robotics and AI*, *5*, 81. https://doi.org/10.3389/frobt.2018.00081

Jones, K. (2019). Trust, distrust, and affective looping. *Philosophical Studies*, *176*(4), 955–968. https://doi.org/10.1007/s11098-018-1221-5

Jung, N., Wranke, C., Hamburger, K., & Knauff, M. (2014). How emotions affect logical reasoning: Evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00570

Jung, S., Sandor, C., Wisniewski, P. J., & Hughes, C. E. (2017). RealME: The influence of body and hand representations on body ownership and presence. *Proceedings of the 5th Symposium on Spatial User Interaction*, 3–11. https://doi.org/10.1145/3131277.3132186

Juvina, I., Collins, M. G., Larue, O., Kennedy, W. G., Visser, E. D., & Melo, C. D. (2019). Toward a Unified Theory of Learned Trust in Interpersonal and Human-Machine Interactions. *ACM Transactions on Interactive Intelligent Systems*, *9*(4), 24:1-24:33. https://doi.org/10.1145/3230735

Kadena, E., & Gupi, M. (2021). Human Factors in Cybersecurity: Risks and Impacts. *Security Science Journal*, *2*(2), 51–64. https://doi.org/10.37458/ssj.2.2.3

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

Kahr, P. K., Rooks, G., Willemsen, M. C., & Snijders, C. C. P. (2023). It Seems Smart, but It Acts Stupid: Development of Trust in AI Advice in a Repeated Legal Decision-Making Task. *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 528–539. https://doi.org/10.1145/3581641.3584058

Kaur, A. (2020). A Link Between Colors and Emotions; A Study of Undergraduate Females. *International Journal of Engineering Research And*, *V9*(09), IJERTV9IS090319. https://doi.org/10.17577/IJERTV9IS090319

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

Khan, A., Sharma, N., & Dixit, S. (2006). *Effect of cognitive load and paradigm on time perception*. *32*(1), 37–42.

Khawaji, A., Chen, F., Zhou, J., & Marcus, N. (2014). Trust and cognitive load in the text-chat environment: The role of mouse movement. *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, 324–327. https://doi.org/10.1145/2686612.2686661

Khawaji, A., Zhou, J., Chen, F., & Marcus, N. (2015). Using Galvanic Skin Response (GSR) to Measure Trust and Cognitive Load in the Text-Chat Environment. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1989–1994. https://doi.org/10.1145/2702613.2732766

Kim, H. J., & Lee, H.-K. (2022). Emotions and Colors in a Design Archiving System: Applying AI Technology for Museums. *Applied Sciences*, *12*(5), 2467. https://doi.org/10.3390/app12052467

Kim, J. (2021). Advertising in the Metaverse: Research Agenda. *Journal of Interactive Advertising*, *21*(3), 141–144. https://doi.org/10.1080/15252019.2021.2001273

Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, *99*(1), 49–65. https://doi.org/10.1016/j.obhdp.2005.07.002

Kim, S.-S., Chung, Y., Perry, M. J., Kawachi, I., & Subramanian, S. V. (2012). Association between Interpersonal Trust, Reciprocity, and Depression in South Korea: A Prospective Analysis. *PLoS ONE*, *7*(1), e30602. https://doi.org/10.1371/journal.pone.0030602

Kim, Y. M., Rhiu, I., & Yun, M. H. (2020). A Systematic Review of a Virtual Reality System from the Perspective of User Experience. *International Journal of Human–Computer Interaction*, *36*(10), 893–910. https://doi.org/10.1080/10447318.2019.1699746

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

Kiser, D. P., Gromer, D., Pauli, P., & Hilger, K. (2022). A virtual reality social conditioned place preference paradigm for humans: Does trait social anxiety affect approach and avoidance of virtual agents? *Frontiers in Virtual Reality, 3*, 916575. https://doi.org/10.3389/frvir.2022.916575

Klingberg, T. (2009). *The overflowing brain: Information overload and the limits of working memory*. Oxford University Press.

Klümper, L., & Sürth, S. (2023). Keep me updated! Social support as a coping strategy to reduce the perceived threat caused by the cognitive availability of COVID-19 relevant information. *Current Psychology*, *42*(8), 6403–6418. https://doi.org/10.1007/s12144-021-01951-w

Koca, Y., & Oriet, C. (2023). From Pictures to the People in Them: Averaging Within-Person Variability Leads to Face Familiarization. *Psychological Science*, *34*(2), 252–264. https://doi.org/10.1177/09567976221131520

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

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

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

Koenig, M. A., & Harris, P. L. (2007). The Basis of Epistemic Trust: Reliable Testimony or Reliable Sources? *Episteme*, *4*(3), 264–284. https://doi.org/10.3366/E1742360007000081

Kolesnichenko, A., McVeigh-Schultz, J., & Isbister, K. (2019). Understanding Emerging Design Practices for Avatar Systems in the Commercial Social VR Ecology. *Proceedings of the 2019 on Designing Interactive Systems Conference*, 241–252. https://doi.org/10.1145/3322276.3322352

Kothgassner, O. D., Goreis, A., Kafka, J. X., Kaufmann, M., Atteneder, K., Beutl, L., Hennig-Fast, K., Hlavacs, H., & Felnhofer, A. (2019). Virtual social support buffers stress response: An experimental comparison of real-life and virtual support prior to a social stressor. *Journal of Behavior Therapy and Experimental Psychiatry, 63*, 57–65. https://doi.org/10.1016/j.jbtep.2018.11.003

Krasovsky, T. (2022). Cognition, Emotion, and Movement in the Context of Rehabilitation. *International Journal of Environmental Research and Public Health*, *19*(21), 14532. https://doi.org/10.3390/ijerph192114532

Kraus, J., Scholz, D., Messner, E.-M., Messner, M., & Baumann, M. (2020). Scared to Trust? – Predicting Trust in Highly Automated Driving by Depressiveness, Negative Self-Evaluations and State Anxiety. *Frontiers in Psychology*, *10*, 2917. https://doi.org/10.3389/fpsyg.2019.02917

Kreps, D. M. (1990). Corporate culture and economic theory. In J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on Positive Political Economy* (pp. 90–143). Cambridge University Press. https://doi.org/10.1017/CBO9780511571657.006

Kugler, T., Ye, B., Motro, D., & Noussair, C. N. (2020). On Trust and Disgust: Evidence From Face Reading and Virtual Reality. *Social Psychological and Personality Science*, *11*(3), 317–325. https://doi.org/10.1177/1948550619856302

Kummer, K. K., Hofhansel, L., Barwitz, C. M., Schardl, A., Prast, J. M., Salti, A., El Rawas, R., & Zernig, G. (2014). Differences in social interaction- vs. Cocaine reward in mouse vs. Rat. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00363

Kupritz, V. W. (2000). PRIVACY MANAGEMENT AT WORK: A CONCEPTUAL MODEL. *Journal of Architectural and Planning Research*, *17*(1), 47–63.

Kwan, L. Y.-Y., Yap, S., & Chiu, C. (2015). Mere exposure affects perceived descriptive norms: Implications for personal preferences and trust. *Organizational Behavior and Human Decision Processes*, *129*, 48–58. https://doi.org/10.1016/j.obhdp.2014.12.002

Kyrlitsias, C., & Michael-Grigoriou, D. (2022). Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey. *Frontiers in Virtual Reality*, *2*, 786665. https://doi.org/10.3389/frvir.2021.786665

Lake, J. (2020). Hey, You Stole My Avatar!: Virtual Reality and Its Risks to Identity Protection. *Emory Law Journal*, *69*(4), 833.

Lane, T. (2017). How does happiness relate to economic behaviour? A review of the literature. *Journal of Behavioral and Experimental Economics*, *68*, 62–78. https://doi.org/10.1016/j.socec.2017.04.001

Lang, B. (2022, May 9). Researchers Show Full-body VR Tracking with Controller-mounted Cameras. *Road to VR*. https://www.roadtovr.com/full-body-vr-tracking-controller-cameras-standalone-headset-research/

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (Technical Report A-8). https://imotions.com/blog/iaps-international-affective-picture-system/

Lange, B., & Pauli, P. (2019). Social anxiety changes the way we move—A social approach-avoidance task in a virtual reality CAVE system. *PLOS ONE*, *14*(12), e0226805. https://doi.org/10.1371/journal.pone.0226805

Langer, M., König, C. J., Back, C., & Hemsing, V. (2023). Trust in Artificial Intelligence: Comparing Trust Processes Between Human and Automated Trustees in Light of Unfair Bias. *Journal of Business and Psychology*, *38*(3), 493–508. https://doi.org/10.1007/s10869-022-09829-9

Latoschik, M. E., & Wienrich, C. (2022). Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. *Frontiers in Virtual Reality*, *3*. https://www.frontiersin.org/articles/10.3389/frvir.2022.694433

Latoschik, M., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017). *The effect of avatar realism in immersive social virtual realities* (p. 10). https://doi.org/10.1145/3139131.3139156

Laux, L. (1981). *Das State-Trait-Angstinventar (STAI): Theoretische Grundlagen und Handanweisung*. Beltz. https://fis.uni-bamberg.de/handle/uniba/26756

Leahy, J. V., & Whited, T. M. (1996). The Effect of Uncertainty on Investment: Some Stylized Facts. *Journal of Money, Credit and Banking*, *28*(1), 64. https://doi.org/10.2307/2077967

Lee, E.-J. (2008). Gender Stereotyping of Computers: Resource Depletion or Reduced Attention? *Journal of Communication*, *58*(2), 301–320. https://doi.org/10.1111/j.1460-2466.2008.00386.x

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50–80.

Lee, W.-S., & Selart, M. (2012). The impact of emotions on trust decisions. In *Handbook on psychology of decision-making: New research* (pp. 1–16). Nova Science Publisher's.

Lee, Y.-H., Xiao, M., & Wells, R. H. (2018). The Effects of Avatars' Age on Older Adults' Self-Disclosure and Trust. *Cyberpsychology Behavior and Social Networking*, *21*(3), 173–178. https://doi.org/10.1089/cyber.2017.0451

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & Emotion*, *14*(4), 473–493. https://doi.org/10.1080/026999300402763

Lester, D., & Gatto, J.-L. (1990). Interpersonal Trust, Depression, and Suicidal Ideation in Teenagers. *Psychological Reports*, *67*(3), 786–786. https://doi.org/10.2466/pr0.1990.67.3.786

Li, S., Ashok, A., Zhang, Y., Xu, C., Lindqvist, J., & Gruteser, M. (2016). Whose move is it anyway? Authenticating smart wearable devices using unique head movement patterns. *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 1–9. https://doi.org/10.1109/PERCOM.2016.7456514

Li, X., Zhang, G., Zhou, C., & Wang, X. (2019). Negative emotional state slows down movement speed: Behavioral and neural evidence. *PeerJ*, *7,* e7591. https://doi.org/10.7717/peerj.7591

Lieberman, A., & Schroeder, J. (2020). Two social lives: How differences between online and offline interaction influence social outcomes. *Current Opinion in Psychology*, *31*, 16–21. https://doi.org/10.1016/j.copsyc.2019.06.022

Liew, T. W., Tan, S.-M., & Ismail, H. (2017). 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.

LimeSurvey Development Team. (2012). *LimeSurvey—The free and open source survey software tool!* [Computer software]. LimeSurvey Development Team. limesurvey.org

Lin, J., Cronje, J., Kathner, I., Pauli, P., & Latoschik, M. E. (2023). Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm. *IEEE Transactions on Visualization and Computer Graphics*, 1–11. https://doi.org/10.1109/TVCG.2023.3247095

Lin, J., & Latoschik, M. E. (2022). Digital body, identity and privacy in social virtual reality: A systematic review. *Frontiers in Virtual Reality*, *3*, 974652. https://doi.org/10.3389/frvir.2022.974652

Lohle, M., & Terrell, S. (2014). Real projects, virtual worlds: Coworkers, their avatars, and the trust conundrum. *Qualitative Report*, *19*.

Lopes, J., Lohan, K., & Hastie, H. (2018). Symptoms of cognitive load in interactions with a dialogue system. *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, 1–5. https://doi.org/10.1145/3279810.3279851

Lount, R. B. (2010). The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of Personality and Social Psychology*, *98*(3), 420–433. https://doi.org/10.1037/a0017344

Lyons, B. R., & Mehta, J. (1997). Contracts, opportunism and trust: Self-interestand social orientation. *Journal of Economics*, *21*(239), 25.

Maalem Lahcen, R. A., Caulkins, B., Mohapatra, R., & Kumar, M. (2020). Review and insight on the behavioral aspects of cybersecurity. *Cybersecurity*, *3*(1), 10. https://doi.org/10.1186/s42400-020-00050-w

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

Mackenzie, B. K. (1948). The importance of contact in determining attitudes toward Negroes. *The Journal of Abnormal and Social Psychology*, *43*(4), 417–441. https://doi.org/10.1037/h0062183

Maloney, D., Zamanifard, S., & Freeman, G. (2020). Anonymity vs. Familiarity: Self-Disclosure and Privacy in Social Virtual Reality. *26th ACM Symposium on Virtual Reality Software and Technology*, 1–9. https://doi.org/10.1145/3385956.3418967

Mancuso, V., Bruni, F., Stramba-Badiale, C., Riva, G., Cipresso, P., & Pedroli, E. (2023). How do emotions elicited in virtual reality affect our memory? A systematic review. *Computers in Human Behavior*, *146*, 107812. https://doi.org/10.1016/j.chb.2023.107812

Marín-Morales, J., Llinares, C., Guixeres, J., & Alcañiz, M. (2020). Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing. *Sensors*, *20*(18), 5163. https://doi.org/10.3390/s20185163

Marsh, S., & Dibben, M. R. (2005). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, *37*(1), 465–498. https://doi.org/10.1002/aris.1440370111

Mattarozzi, K., Todorov, A., Marzocchi, M., Vicari, A., & Russo, P. M. (2015). Effects of Gender and Personality on First Impression. *PLOS ONE*, *10*(9), e0135529. https://doi.org/10.1371/journal.pone.0135529

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320. https://doi.org/10.1037/0022-0663.90.2.312

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the Effect of Workload on Automation Use for Younger and Older Adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*(6), 672–686. https://doi.org/10.1177/0018720811421909

McCambridge, J., Bruin, M. de, & Witton, J. (2012). The Effects of Demand Characteristics on Research Participant Behaviours in Non-Laboratory Settings: A Systematic Review. *PLOS ONE, 7*(6), e39116. https://doi.org/10.1371/journal.pone.0039116

McKendrick, G., & Graziane, N. M. (2020). Drug-Induced Conditioned Place Preference and Its Practical Use in Substance Use Disorder Research. *Frontiers in Behavioral Neuroscience*, *14*, 582147. https://doi.org/10.3389/fnbeh.2020.582147

McVeigh-Schultz, J., Márquez Segura, E., Merrill, N., & Isbister, K. (2018). What's It Mean to "Be Social" in VR?: Mapping the Social VR Design Ecology. *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems,* 289–294. https://doi.org/10.1145/3197391.3205451

Melnick, K. (2022, February 17). VR Social Platform VRChat Adds Face Tracking For Avatars. *VRScout*. https://vrscout.com/news/vr-social-platform-vrchat-adds-face-tracking-for-avatars/

Menges, U., Hielscher, J., Buckmann, A., Kluge, A., Sasse, M. A., & Verret, I. (2022). Why IT Security Needs Therapy. In S. Katsikas, C. Lambrinoudakis, N. Cuppens, J. Mylopoulos, C. Kalloniatis, W. Meng, S. Furnell, F. Pallas, J. Pohle, M. A. Sasse, H. Abie, S. Ranise, L. Verderame, E. Cambiaso, J. Maestre Vidal, & M. A. Sotelo Monge (Eds.), *Computer Security. ESORICS 2021 International Workshops* (Vol. 13106, pp. 335–356). Springer International Publishing. https://doi.org/10.1007/978-3-030-95484-0_20

Mesurado, B., Resett, S., Tezón, M., & Vanney, C. E. (2021). Do Positive Emotions Make You More Prosocial? Direct and Indirect Effects of an Intervention Program on Prosociality in Colombian Adolescents During Social Isolation Due to COVID-19. *Frontiers in Psychology, 12,* 710037. https://doi.org/10.3389/fpsyg.2021.710037

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

Mills, C., & D'Mello, S. (2014). On the Validity of the Autobiographical Emotional Memory Task for Emotion Induction. *PLOS ONE, 9*(4), 10.

Montañez, R., Golob, E., & Xu, S. (2020). Human Cognition Through the Lens of Social Engineering Cyberattacks. *Frontiers in Psychology*, *11*, 1755. https://doi.org/10.3389/fpsyg.2020.01755

Montoro, A., Shih, P.-C., Román, M., & Martínez-Molina, A. (2014). Adaptación al castellano de la Escala de Confianza General de Yamagishi. *Anales de Psicología*, *30*(1), 302–307. https://doi.org/10.6018/analesps.30.1.122471

Moorman, C., Deshpande, R., & Zaltman, G. (1993). Factors Affecting Trust in Market Research Relationships. *Journal of Marketing*, *57*(1), 81. https://doi.org/10.2307/1252059

Moradinezhad, R., & Solovey, E. T. (2021). Investigating Trust in Interaction with Inconsistent Embodied Virtual Agents. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-021-00747-z

Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, *58*(3), 20. https://doi.org/10.2307/1252308

Morriss, J., Tupitsa, E., Dodd, H. F., & Hirsch, C. R. (2022). Uncertainty Makes Me Emotional: Uncertainty as an Elicitor and Modulator of Emotional States. *Frontiers in Psychology*, *13*, 777025. https://doi.org/10.3389/fpsyg.2022.777025

Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates and scenarios. *Computers & Security*, *59*, 186–209. https://doi.org/10.1016/j.cose.2016.03.004

Mulken, S. van, André, E., & Müller, J. (1999). An empirical study on the trustworthiness of life-like interface agents. *Human-Computer Interaction (Proceedings of Hci-International 1999), 152–156. Mahwah*, 152–156.

Müller, S., Schwehm, S., Hewig, J., & Paelecke, M. (2020). Influence of Trust-Enhancing Signals on Signal Processing in B2C E-Commerce. *AMCIS 2020 Proceedings*. https://aisel.aisnet.org/amcis2020/sig_hci/sig_hci/6

Myers, C. D., & Tingley, D. (2016). The Influence of Emotion on Trust. *Political Analysis*, *24*(4), 492–500. https://doi.org/10.1093/pan/mpw026

Mystakidis, S., Berki, E., & Valtanen, J.-P. (2021). Deep and Meaningful E-Learning with Social Virtual Reality Environments in Higher Education: A Systematic Literature Review. *Applied Sciences*, *11*(5), Article 5. https://doi.org/10.3390/app11052412

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Conference Companion on Human Factors in Computing Systems - CHI '94*, 73–78. https://doi.org/10.1145/259963.260288

Nass, C., Steuer, J., Tauber, E. R., & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. *CHI '93*. https://doi.org/10.1145/259964.260137

Nass, Moon, Fogg, Reeves, & Dryer. (1995). *Can Computer Personalities Be Human Personalities? 43*(2), 223-239.

Nordin Forsberg, B., & Kirchner, K. (2021). The Perception of Avatars in Virtual Reality During Professional Meetings. In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), *HCI International 2021—Posters* (pp. 290–294). Springer International Publishing. https://doi.org/10.1007/978-3-030-78642-7_39

Normand, V., Babski, C., Benford, S., Bullock, A., Carion, S., Chrysanthou, Y., Farcet, N., Frécon, E., Harvey, J., Kuijpers, N., Magnenat-Thalmann, N., Raupp-Musse, S., Rodden, T., Slater, M., Smith, G., Steed, A., Thalmann, D., Tromp, J., Usoh, M., … Kladias, N. (1999). The COVEN Project: Exploring Applicative, Technical, and Usage Dimensions of Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *8*(2), 218–236. https://doi.org/10.1162/105474699566189

O'Brolcháin, F., Jacquemard, T., Monaghan, D., O'Connor, N., Novitzky, P., & Gordijn, B. (2016). The Convergence of Virtual Reality and Social Networks: Threats to Privacy and Autonomy. *Science and Engineering Ethics*, *22*(1), 1–29. https://doi.org/10.1007/s11948-014-9621-1

Oculus Lipsync for Unity Development: Unity | Oculus Developers. (n.d.). Retrieved May 26, 2023, from https://developer.oculus.com/documentation/unity/audio-ovrlipsync-unity/

Olade, I., Fleming, C., & Liang, H.-N. (2020). BioMove: Biometric User Identification from Human Kinesiological Movements for Virtual Reality Systems. *Sensors*, *20*(10), 2944. https://doi.org/10.3390/s20102944

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. https://doi.org/10.1007/s10919-009-0082-1

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oyebode, B., & Nicholls, N. (2020). Social norms as anchor points for trust. *The Social Science Journal*, 1–11. https://doi.org/10.1080/03623319.2020.1753429

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Pan, Y., & Steed, A. (2016). A Comparison of Avatar-, Video-, and Robot-Mediated Interaction on Users’ Trust in Expertise. *Frontiers in Robotics and AI*, *3*. https://doi.org/10.3389/frobt.2016.00012

Pan, Y., & Steed, A. (2017). The impact of self-avatars on trust and collaboration in shared virtual environments. *Plos One*, *12*(12), https://doi.org/10.1371/journal.pone.0189078

Pasquini, E., Corriveau, K., Koenig, M., & Harris, P. (2007). Preschoolers Monitor the Relative Accuracy of Informants. *Developmental Psychology*, *43*, 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Peña, J., & Yoo, S.-C. (2014). Under Pressure: Avatar Appearance and Cognitive Load Effects on Attitudes, Trustworthiness, Bidding, and Interpersonal Distance in a Virtual Store. *Presence: Teleoperators and Virtual Environments*, *23*(1), 18–32. https://doi.org/10.1162/PRES_a_00166

Peskin, M., & Newell, F. N. (2004). Familiarity Breeds Attraction: Effects of Exposure on the Attractiveness of Typical and Distinctive Faces. *Perception*, *33*(2), 147–157. https://doi.org/10.1068/p5028

Pfaller, M., Kroczek, L. O. H., Lange, B., Fülöp, R., Müller, M., & Mühlberger, A. (2021). Social Presence as a Moderator of the Effect of Agent Behavior on Emotional Experience in Social Interactions in Virtual Reality. *Frontiers in Virtual Reality*, *2*, 741138. https://doi.org/10.3389/frvir.2021.741138

Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive load theory*. Cambridge University Press.

PriMa – Privacy Matters. (2018, April 19). *PriMa*. https://www.prima-itn.eu/

Principe, C. P., & Langlois, J. H. (2013). Children and adults use attractiveness as a social cue in real people and avatars. *Journal of Experimental Child Psychology*, *115*(3), 590–597. https://doi.org/10.1016/j.jecp.2012.12.002

Radiah, R., Roth, D., Alt, F., & Abdelrahman, Y. (2023). The Influence of Avatar Personalization on Emotions in VR. *Multimodal Technologies and Interaction*, *7*(4), 38. https://doi.org/10.3390/mti7040038

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rau, H. A. (2012). Trust and trustworthiness—A survey of gender differences. In *Psychology of gender differences* (pp. 205–224). Nova Science Publishers.

Rebelo, F., Noriega, P., Duarte, E., & Soares, M. (2012). Using Virtual Reality to Assess User Experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(6), 964–982. https://doi.org/10.1177/0018720812465006

Reeb-Sutherland, B. C., Fifer, W. P., Byrd, D. L., Hammock, E. A. D., Levitt, P., & Fox, N. A. (2011). One-month-old human infants learn about the social world while they sleep: Infants learn while they sleep. *Developmental Science*, *14*(5), 1134–1141. https://doi.org/10.1111/j.1467-7687.2011.01062.x

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

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable Facial Configurations Affect Strategic Choices in Trust Games with or without Information about Past Behavior. *PLoS ONE*, *7*(3), e34293. https://doi.org/10.1371/journal.pone.0034293

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

Robinson, J. (2014). Faith in People. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 2151–2152). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_989

Rogers, C. E., Witt, A. W., Solomon, A. D., & Venkatasubramanian, K. K. (2015). *An Approach for User Identification for Head-Mounted Displays*. 4.

Rosenberg, M. (1956). Misanthropy and Political Ideology. *American Sociological Review*, *21*(6), 690–695. https://doi.org/10.2307/2088419

Rosenberg, R. S., Baughman, S. L., & Bailenson, J. N. (2013). Virtual Superheroes: Using Superpowers in Virtual Reality to Encourage Prosocial Behavior. *PLoS ONE*, *8*(1), e55003. https://doi.org/10.1371/journal.pone.0055003

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

Roth, D., Kleinbeck, C., Feigl, T., Mutschler, C., & Latoschik, M. (2018). *Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities*. https://doi.org/10.1109/VR.2018.8447550

Roth, D., & Latoschik, M. E. (2020). Construction of the Virtual Embodiment Questionnaire (VEQ). *IEEE Transactions on Visualization and Computer Graphics*, *26*(12), 3546–3556. https://doi.org/10.1109/TVCG.2020.3023603

Roth, D., Waldow, K., Latoschik, M. E., Fuhrmann, A., & Bente, G. (2017). Socially immersive avatar-based communication. *2017 IEEE Virtual Reality (VR)*, 259–260. https://doi.org/10.1109/VR.2017.7892275

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

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. *Presence: Teleoperators and Virtual Environments*, *11*(4), 335–351. https://doi.org/10.1162/105474602760204264

Rydval, O. (2012). The Causal Effect of Cognitive Abilities on Economic Behavior: Evidence from a Forecasting Task with Varying Cognitive Load. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2046942

Sadeh, N., & Bredemeier, K. (2021). Engaging in Risky and Impulsive Behaviors to Alleviate Distress Mediates Associations Between Intolerance of Uncertainty and Externalizing Psychopathology. *Journal of Personality Disorders*, *35*(3), 393–408. https://doi.org/10.1521/pedi_2019_33_456

Salanitri, D., Lawson, G., & Waterfield, B. (2016). The Relationship Between Presence and Trust in Virtual Reality. *Proceedings of the European Conference on Cognitive Ergonomics*, 1–4. https://doi.org/10.1145/2970930.2970947

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

Sandstrom, G. M., Boothby, E. J., & Cooney, G. (2022). Talking to strangers: A week-long intervention reduces psychological barriers to social connection. *Journal of*

Experimental Social Psychology, *102*, 104356. https://doi.org/10.1016/j.jesp.2022.104356

Sansoni, M., Scarzello, G., Serino, S., Groff, E., & Riva, G. (2022). Mitigating negative emotions through virtual reality and embodiment. *Frontiers in Human Neuroscience*, *16*, 916227. https://doi.org/10.3389/fnhum.2022.916227

Scavarelli, A., Arya, A., & Teather, R. J. (2021). Virtual reality and augmented reality in social learning spaces: A literature review. *Virtual Reality*, *25*(1), 257–277. https://doi.org/10.1007/s10055-020-00444-8

Schubert, T. W. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, *15*(2), 69–71. https://doi.org/10.1026//1617-6383.15.2.69

Schwarz, N. (2012). Feelings-as-Information Theory. In P. Van Lange, A. Kruglanski, & E. Higgins, *Handbook of Theories of Social Psychology: Volume 1* (pp. 289–308). SAGE Publications Ltd. https://doi.org/10.4135/9781446249215.n15

Serva, M. A., Fuller, M. A., & Mayer, R. C. (2005). The Reciprocal Nature of Trust: A Longitudinal Study of Interacting Teams. *Journal of Organizational Behavior*, *26*(6), 625–648.

Shiv, B., & Fedorikhin, A. (1999). Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making. *Journal of Consumer Research*, *26*(3), 278–292. https://doi.org/10.1086/209563

Siegrist, M., Gutscher, H., & Earle, T. C. (2005). Perception of risk: The influence of general trust, and general confidence. *Journal of Risk Research*, *8*(2), 145–156. https://doi.org/10.1080/1366987032000105315

Siehl, S., Kammler-Sücker, K., Guldner, S., Zohair, R., Janvier, Y., & Nees, F. (2022). *To trust or not to Trust? Face and voice modulation of virtual avatars* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/b4ka7

Silic, M., & Cyr, D. (2016). Colour Arousal Effect on Users' Decision-Making Processes in the Warning Message Context. In F. F.-H. Nah & C.-H. Tan (Eds.), *HCI in Business, Government, and Organizations: Information Systems* (Vol. 9752, pp. 99–109). Springer International Publishing. https://doi.org/10.1007/978-3-319-39399-5_10

Slater, M. (2008). Towards a digital body: The virtual arm illusion. *Frontiers in Human Neuroscience*, *2*. https://doi.org/10.3389/neuro.09.006.2008

Slater, M., & Usoh, M. (1999). Body Centred Interaction in Immersive Virtual Environments. *Artificial Life and Virtual Reality*.

Sluganovic, I., Roeschlin, M., Rasmussen, K. B., & Martinovic, I. (2016). Using Reflexive Eye Movements for Fast Challenge-Response Authentication. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 1056–1067. https://doi.org/10.1145/2976749.2978311

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What Is Typical Is Good: The Influence of Face Typicality on Perceived Trustworthiness. *Psychological Science*, *26*(1), 39–47. https://doi.org/10.1177/0956797614554955

Spering, M., Wagener, D., & Funke, J. (2005). The role of emotions in complex problem solving. *Cognition & Emotion*, *19*(8), 1252–1261. https://doi.org/10.1080/02699930500304886

Surprenant, A. (2012). Measuring Trust In Virtual Worlds: Avatar-mediated Self-disclosure. *Electronic Theses and Dissertations*. https://stars.library.ucf.edu/etd/2157

Survey Research Center. (1969). Election study. Ann Arbor: Inter-University Consortium for Political Research. *University of Michigan.* https://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Self_Measures_for_Trust_TRUST_IN_PEOPLE.pdf

Susindar, S., Sadeghi, M., Huntington, L., Singer, A., & Ferris, T. K. (2019). The Feeling is Real: Emotion Elicitation in Virtual Reality. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 252–256. https://doi.org/10.1177/1071181319631509

Swann, W. B., Hixon, J. G., Stein-Seroussi, A., & Gilbert, D. T. (1990). The fleeting gleam of praise: Cognitive processes underlying behavioral reactions to self-relevant feedback. *Journal of Personality and Social Psychology*, *59*(1), 17–26. https://doi.org/10.1037//0022-3514.59.1.17

Swap, W. C. (1977). Interpersonal Attraction and Repeated Exposure to Rewarders and Punishers. *Personality and Social Psychology Bulletin*, *3*(2), 248–251. https://doi.org/10.1177/014616727700300219

Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Thalmann, D. (2001). The Role of Virtual Humans in Virtual Environment Technology and Interfaces. In R. A. Earnshaw, R. A. Guedj, A. van Dam, & J. A. Vince (Eds.), *Frontiers of Human-Centered Computing, Online Communities and Virtual Environments* (pp. 27–38). Springer. https://doi.org/10.1007/978-1-4471-0259-5_3

The Jamovi project. (2021). *Jamovi* (1.8.4.0) [Computer software]. https://www.jamovi.org.

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality*, *50*(1), 61–65. https://doi.org/10.1016/j.jrp.2014.03.006

Thielmann, I., & Hilbig, B. E. (2015). Trust: An Integrative Review from a Person–Situation Perspective. *Review of General Psychology*, *19*(3), 249–277. https://doi.org/10.1037/gpr0000046

Tislar, C., Sterkenburg, J., Zhang, W., & Jeon, M. (2014). How Emotions Influence Trust in Online Transactions Using New Technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1531–1535. https://doi.org/10.1177/1541931214581319

Tomlinson, E. C., & Mryer, R. C. (2009). The Role Of Causal Attribution Dimensions In Trust Repair. *Academy of Management Review*, *34*(1), 85–104. https://doi.org/10.5465/amr.2009.35713291

Torre, I. (2017). *The impact of voice on trust attributions* [Thesis, University of Plymouth]. https://pearl.plymouth.ac.uk/handle/10026.1/9858

Trabucco, J. T., Rottigni, A., Cavallo, M., Bailey, D., Patton, J., & Marai, G. E. (2019). User perspective and higher cognitive task-loads influence movement and performance in immersive training environments. *BMC Biomedical Engineering*, *1*(1), 21. https://doi.org/10.1186/s42490-019-0021-0

Tripp, J., McKnight, H., & Lankton, N. K. (2011). *Degrees of Humanness in Technology: What Type of Trust Matters?*

Tsankova, E., Aubrey, A., Krumhuber, E., Möllering, G., Kappas, A., Marshall, D., & Rosin, P. (2013). *Facial and Vocal Cues in Perceptions of Trustworthiness* (Vol. 7729). https://doi.org/10.1007/978-3-642-37484-5_26

Unity Engine. (n.d.). *Unity Engine* (2020.3.14f) [Computer software]. Unity Software Inc.

Uzzi, B., & Gillespie, J. J. (2002). Knowledge spillover in corporate financing networks: Embeddedness and the firm’s debt performance. *Strategic Management Journal*, *23*(7), 595–618. https://doi.org/10.1002/smj.241

Van Kleef, G. A., Cheshin, A., Fischer, A. H., & Schneider, I. K. (2016). Editorial: The Social Nature of Emotions. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00896

Vugt, H. C. V., Bailenson, J. N., Hoorn, J. F., & Konijn, E. A. (2010). Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction*, *17*(2), 1–27. https://doi.org/10.1145/1746259.1746261

Wallace, B., Ross, A., & Davies, J. (2003). *Information processing models: Benefits and limitations*. Taylor & Francis. https://doi.org/10.1201/b12800

Wang, W. (2020). Difference between the Real World and Virtual World. *IS4SI 2019 Summit*, 35. https://doi.org/10.3390/proceedings2020047035

Wang, Z., Zhu, H., & Sun, L. (2021). Social Engineering in Cybersecurity: Effect Mechanisms, Human Vulnerabilities and Attack Methods. *IEEE Access*, *9*, 11895–11910. https://doi.org/10.1109/ACCESS.2021.3051633

Ward, A., & Mann, T. (2000). Don't mind if I do: Disinhibited eating under cognitive load. *Journal of Personality and Social Psychology*, *78*(4), 753–763. https://doi.org/10.1037//0022-3514.78.4.753

Wargo, E. (2006). How Many Seconds to a First Impression? *APS Observer*, *19*. https://www.psychologicalscience.org/observer/how-many-seconds-to-a-first-impression

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wheeless, L. R., & Grotz, J. (1977). THE MEASUREMENT OF TRUST AND ITS RELATIONSHIP TO SELF-DISCLOSURE. *Human Communication Research*, *3*(3), 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x

Willis, J., & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilner, D. M., Walkley, R. P., & Cook, S. W. (1952). Residential Proximity and Intergroup Relations in Public Housing Projects. *Journal of Social Issues*, *8*(1), 45–69. https://doi.org/10.1111/j.1540-4560.1952.tb01593.x

Wingler, D., Joseph, A., Bayramzadeh, S., & Robb, A. (2020). Using Virtual Reality to Compare Design Alternatives Using Subjective and Objective Evaluation Methods. *HERD: Health Environments Research & Design Journal*, *13*(1), 129–144. https://doi.org/10.1177/1937586719851266

Wrightsman, L. S. (1964). Measurement of Philosophies of Human Nature. *Psychological Reports*, *14*(3), 743–751. https://doi.org/10.2466/pr0.1964.14.3.743

Wu, Y., Hall, A. S. M., Siehl, S., Grafman, J., & Krueger, F. (2020). Neural Signatures of Gender Differences in Interpersonal Trust. *Frontiers in Human Neuroscience, 14*, 225. https://doi.org/10.3389/fnhum.2020.00225

Yamagishi, T. (2011). *Trust: The Evolutionary Game of Mind and Society* (2011th ed.). Springer.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 129–166. https://doi.org/10.1007/BF02249397

Yang, W. (2018). *Affective auditory stimulus database: An expanded version of the International Affective Digitized Sounds (IADS-E)*. 15.

Yaremych, H. E., & Persky, S. (2019). Tracing physical behavior in virtual reality: A narrative review of applications to social psychology. *Journal of Experimental Social Psychology*, *85*, 103845. https://doi.org/10.1016/j.jesp.2019.103845

Yildiz, M. N., Fazal, M., Ahn, M., Feirsen, R., & Ozdemir, S. (Eds.). (2019). *Handbook of Research on Media Literacy Research and Applications Across Disciplines:* IGI Global. https://doi.org/10.4018/978-1-5225-9261-7

Yu, Z., Liang, H.-N., Fleming, C., & Man, K. L. (2016). An exploration of usable authentication mechanisms for virtual reality systems. *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 458–460. https://doi.org/10.1109/APCCAS.2016.7804002

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zanbaka, C. A., Ulinski, A. C., Goolkasian, P., & Hodges, L. F. (2007). Social responses to virtual humans: Implications for future interface design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1561–1570. https://doi.org/10.1145/1240624.1240861

Zebrowitz, L. A., Bronstad, P. M., & Lee, H. K. (2007). The Contribution of Face Familiarity to Ingroup Favoritism and Stereotyping. *Social Cognition*, *25*(2), 306–338. https://doi.org/10.1521/soco.2007.25.2.306

Zhou, J., Arshad, S. Z., Luo, S., & Chen, F. (2017). Effects of Uncertainty and Cognitive Load on User Trust in Predictive Decision Making. In R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, & M. Winckler (Eds.), *Human-Computer*

Interaction – INTERACT 2017 (pp. 23–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-68059-0_2

Appendix

Appendix A

Summary of the measures of trust

Subjective Measurements

- Philosophies of Human Nature Scale (PHN) Trustworthiness subscale
- Interpersonal Trust Scale (ITS) (Rotter, 1967)
- Trust in People Scale (Survey Research Center, 1969)
- Specific interpersonal trust scale (SITS)(Johnson-George & Swap, 1982)
- General Trust Scale (GTS) (Yamagishi & Yamagishi, 1994c)
- The Propensity to Trust Survey (PTS) (Evans & Revelle, 2008)
- The short scale interpersonal trust (KUSIV3) (Beierlein et al., 2012)
- The Social Trust Scale of the European Social Survey (ESS) (Breyer, 2015)
- World Values Survey (Haerpfer et al., 2020)

Objective Measurements

Behavioural measurements

- Lending game (Camerer & Weigelt, 1988)
- Trust game- binary choice (Kreps, 1990)
- Gift Exchange Game (Fehr et al., 1993)
- Investor game (Berg et al., 1995)
- Trading game (Lyons & Mehta, 1997)
- Envelop drop (Glaeser et al., 2000)
- Experimental maze task (Hale et al., 2018)
- Multi-arm trust game (Juvina et al., 2019)

Overview of subjective trust measures

A variety of subjective measures of trust were proposed. In this review only the measures that focus on interpersonal trust were included. Table 13 lists selected questionnaires and scales that were proposed to measure interpersonal trust. This review of the subjective interpersonal trust measurements starts with a general measurement of interpersonal trust and is followed by more specific interpersonal trust measurements.

Table 13: Summary of subjective trust measurements in experimental studies

Authors	Subjective measure of Trust	Scale description and focus	Number of items	Languages	Conceptualization of interpersonal trust
Interpersonal trust as part of human nature					
Rosenberg (1956)	Faith in People	Respondents global attitude towards human nature and political ideology	5 items	English	-
Wrightsman (1964)	Philosophies of Human Nature Scale (PHN)_ Trustworthiness subscale	Respondents indicate how they perceive other people, more specifically to which extent they regard them as moral, reliable, and honest.	8 items	English	The extent to which people are seen as moral, honest, and reliable.
Generalized interpersonal trust					
Rotter (1967)	Interpersonal Trust Scale (ITS)	This scale measures the generalized expectation of being able to rely on the words and promises of interaction partners, either based on verbal or written utterances. Used with experimental trust task, such as the Prisoners dilemma.	25 items	German, English	An expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon
Yamagishi & Yamagishi (1994)	General Trust Scale (GTS)	This self-reported tool measures predictable trusting behaviour, and it has proven useful in cross-cultural research.	6 items	English, Polish, Spanish	Trust is the expectation of goodwill and benign intent.
	Interpersonal Trust Short Scale (KUSIV3)	The KUSIV3 is a short scale for capturing the psychological trait of interpersonal trust in social science surveys.	Reduced the Interpersonal trust scale from (Rotter, 1967) to 3 items	German, English	An expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon
Breyer (2015)	The European Social Survey (ESS)_ The Social Trust subscale	Measures to what extent respondents trust and expect fairness from others.	3 items	English and 26 further languages	-
Haerpfer et al. (2020)	World Values Survey	Measurement of individual beliefs. Investigate respondents' general thoughts about others, whether they are worth trusting or whether it is better to be careful.	Single item scale	English	-
Individual difference in interpersonal trust					
Johnson-George & Swap (1982)	Specific interpersonal trust scale (SITS)	Measurement of the varieties of interpersonal trust held by one individual for a specific other person. A male and female version available taking individual gender effects on trustworthiness into consideration.	50 items	English	-
Evans & Revelle (2008)	The Propensity to Trust Survey (PTS)	Used to predict behaviour in the Investment Game (Berg et al., 1995), an experimental trust task. According to Evans & Revelle (2008) the PTS is preferable to the general Big Five personality measure for predicting trusting behaviour. Individual differences in trust can be observed.	26 items	English	Conceptualizes trust as an persisting trait instead of a momentary state

Review of subjective measurements of interpersonal trust in experimental research

The *"Faith in People" scale* was developed in the field of politics (Rosenberg, 1956). This scale focus on respondents' global attitude towards human nature and political ideology, more specifically the factors (interpersonal relationships, group affiliations and personal characteristics) that influence political ideology and political attitude (Rosenberg, 1956). The "Faith in People" scale has a coefficient of reproducibility of 0.92 (Rosenberg, 1956). However, this measure does not focus on the interpersonal trust directly. Wrightsman (1964) criticized the "Faith in People" scales simplistic approach to human nature and suggest that human nature has independent dimensions, including trust. The *"Philosophies of Human Nature Scale" (PHN)* was developed by Wrightsman in 1964 and finds its origin in the fields of theology, social science and philosophy (Wrightsman, 1964). The PHN measures how people perceive each other and assess people's philosophies of human nature. This scale consists of six bipolar subscales (Trustworthiness, Altruism, Independence, Strength of Will and Rationality, Complexity, and Variability). Each component is measured with a Likert scale. After two-item analyses, 14 items on each of six subscales, were constructed (Wrightsman, 1964). Only the trustworthiness subscale is relevant for this review. The trustworthiness subscale is part of the PHN and investigates.

During the development of the PHN, it was administered to 177 undergraduate students from 3 colleges in the South and the Midwest of America (Wrightsman, 1964). The PHN contributes towards predictions of sex differences, self-ideal discrepancies, religious differences, and evaluations of one's instructor. These subscales have adequate internal consistency and strong consistency over time (Wrightsman, 1964). The split-half reliability coefficient for the trustworthiness subscale is 0.74 and the trustworthiness subscale (8 items) is available in English (Wrightsman, 1964). The PHN has a positive correlation with the "Faith in People" scale ranging from 0.39 and 0.75. This is anticipated, as both scales attempt to measure the goodness, worthiness, and improvability of human nature (Robinson, 2014; Wrightsman, 1964). Building on these measures that regard trust in relation to human nature, is the "Social trust subscale" of the European Social Survey (ESS) established in the field of social psychology, politics and media and communication studies (Breyer, 2015). This longitudinal study monitors changing attitudes and values across Europe since 2001. *"The Social trust subscale"* of the ESS was applied since the beginning of the ESS in 2001 and consists of three items measuring misanthropy on an international level (Breyer, 2015). During the development of the social trust subscale, it was administered to 54,673 participants (25,214 males and 29,442 females) from 29 countries with a mean age of 47.08 years and a standard deviation of 18.43 years (Breyer, 2015).The social trust scale measures interpersonal trust, by investigating to what extent respondents trust and expect fairness from other people. This scale was originally developed in British English and then translated to 26 further languages (Breyer, 2015). The Social trust subscale has a Cronbach's alpha ranging from 0.52 to 0.85 (Breyer, 2015) .

Rotter developed the first measuring instrument for recording a generalized interpersonal trust personality trait with the *"Interpersonal Trust scale" (ITS)* (Rotter, 1967). The ITS has its foundation in the field of social and developmental psychology. Rotter postulated that trust, assessed by his scale, is most accurate in highly uncertain, novel, or unstructured situations, where one's generalized expectancy is all one can rely on (Rotter, 1967). The social learning theory and developmental psychology laid the foundation for the development of the Interpersonal Trust Scale (Dai et al., 2020; Rotter, 1967). This scale investigates the generalized expectation of being able to rely on the words and promises (either verbal or written utterances) of others (Rotter, 1967). This subjective measurement scale is often combined with experimental tasks, such as the Prisoners dilemma or lending game (Camerer & Weigelt, 1988) that investigate interpersonal trust (Alós-Ferrer & Farolfi, 2019). This scale consists of 25 items and is available in English and in German and was first evaluated with 547 participants (Rotter, 1967). The ITS asks participants to indicate their trust in a variety of social interaction partners (for example, parents, teachers, or physicians). Higher scores indicate higher interpersonal trust in social interaction partners (Rotter, 1967). This scale has high internal consistency and the split-half reliability coefficient is 0.76 (Rotter, 1967). The ITS is often applied in experimental research and was translated from English to German and Chinese (Beierlein, Kemper, & Rammstedt, Beatrice, 2012; Dai et al., 2020).

In 2012 the Interpersonal Trust Scale of Rotter was adapted and reduced to a 3-item scale and was developed in German. This shorted scale is known as the *"KUSIV3"* (English and German) and is a highly economical scale for capturing the psychological trait of interpersonal trust in social science surveys (Beierlein et al., 2012). In this 3 items scale the KUSIV3 measures general trust in strangers or close acquaintances (Beierlein et al., 2012). Respondents indicate general trust on a 5-point rating scale ranging from "strongly disagree" (1) to "strongly agree" (5). In the development of the KUSIV3, the scale was administered to the general German-speaking population over 18 years of age (Beierlein et al., 2012). The estimator of the reliability for this scale was determined based on loads of the three items on a common factor. The KUSIV3 has a reliability coefficient of 0.85 and has a strong correlation with the ITS and the SOEP- trust scale (Beierlein et al., 2012). This KUSIV3 scale are rooted in the fields of social and developmental psychology.

Another measure of general trust is the *"General Trust Scale" (GTS)* that determines behavioural predictions cross-culturally, between homogeneous groups with high and low trustors (Yamagishi & Yamagishi, 1994). The GTS finds its theoretical origin in behavioural economics and social psychology. The Japanese researchers etymologically deconstructed the word "trust" in their native language to provide more transparency in its meaning. Yamagishi & Yamagishi (1994) suggest that this scale is an alternative to Rotter's interpersonal trust scale as well as the *"World Values Survey",* which is a one-item scale (asking whether people can generally be trusted or that one should rather be careful) (Haerpfer et al., 2020; Jasielska et al., 2019). The use of GTS is effective when seeking to understand the general level of trust. According to the authors, the GTS is appropriate in measuring cross-cultural variations of trust in experimental and behavioural economic research (Yamagishi & Yamagishi, 1994). Despite the variety of trust definitions, when authors focus on the propensity to trust most agree on its description: it is the general expectation that others will behave benevolently and with goodwill (Rotter, 1967; Yamagishi & Yamagishi, 1994). During the development of the general trust scale, 246 American students and 928 Japanese students were recruited for a two-phase development, both in English and in Japanese (Yamagishi & Yamagishi, 1994).

There are strong positive correlations between the following interpersonal trust scales: general trust scale (Yamagishi & Yamagishi, 1994), interpersonal trust scale (Rotter, 1967), the Philosophies of human nature scale (Wrightsman, 1964). The general trust scale consists of 6-items with its internal reliability ranging from 0.70 to 0.78 and a strong predictive validity (Yamagishi & Yamagishi, 1994). The GTS has been applied in experimental tasks presenting the social dilemma and used to explore the role of trust in organizational settings (Montoro et al., 2014).

Interpersonal trust is a complex, multi-faceted phenomenon that is different for different social interaction partners in different settings (e.g. organizational or romantic) (Yamagishi & Yamagishi, 1994). For example, interpersonal trust between romantic partners is different than interpersonal trust between strangers. Interpersonal trust has been shown to involve various components that can be measured independently and fluctuate as a function of the precise nature and meaning of the target person's behaviour (Johnson-George & Swap, 1982). The interpersonal trust scale, the philosophies of human nature scale, and the trust in people scale were developed to measure the general inclinations of a person to trust another (Johnson-George & Swap, 1982; Rotter, 1967; Wrightsman, 1964). These scales can make general predictions of a person's trust inclination towards social interaction partners in general. However, they are not able to accurately determine trust inclination under specific circumstances. This point inspired Johnson-George and Swap to develop the *"Specific interpersonal trust scale" (SITS)* with a focus on the varieties of interpersonal trust, held by one individual for another (Johnson-George & Swap, 1982). The "specific trust scale" consists of 50 items that describe hypothetical situations. These items fall into four categories: 1) trusting another with one's material possessions, 2) a belief in the other's dependability or reliability, 3) trusting another with personal confidences, and 4) trusting another with one's physical safety (Johnson-George & Swap, 1982). This scale was developed by its application to a student's sample of 405 participants (180 male and 225 female) with a trust manipulation. During the trust manipulation participants were asked to wait in a soundproof room for their interaction partner and complete eight coordination problems, the interaction partner would either be on time (contributing towards trust) or 10 minutes late (impacting trust negatively) for the experiment (Johnson-George & Swap, 1982). Male and female results were separately analyzed utilizing factor analysis to produce a male and a female specific interpersonal trust scale. The 19-item *"SITS-M"* (male scale) and a 13-item *"SITS-F"* (female scale) was constructed. This scale suggests that trust can be measured separately from other positive interpersonal qualities with reliable and valid items. The reliability coefficient for both SITS-M and SITS-F is 0.71. The SITS is suitable for experimental use in clinical and laboratory settings to determine the levels of trust toward people who are typically viewed as trustworthy (for example parents, spouses, or therapists) (Johnson-George & Swap, 1982).

The *"Propensity to trust" (PTS)* scale is concerned with the individual differences in interpersonal trust and predicting trust behaviour. It is widely applied in behavioural economics research (Camerer, 2010; Evans & Revelle, 2008). In the development of the PTS it was administered to 8183 participants with 73% female participants, the mean age was 28 (Evans & Revelle, 2008). Evans and Revelle (2008) claim trust to be a stable trait rather than a temporary state, they challenge the finding that trust is purely situational and provides evidence for trust as a personality trait. High trust scores on the PTS positively correlate with the dimensions agreeableness ($r = 0.27$) and extraversion ($r = 0.66$) of the Big Five Inventory (Evans & Revelle, 2008). The investment game (of Berg et al., (1995) was used as an experimental task to validate the inventory of the PTS. The PTS is statistically reliable with a reliability coefficient of 0.75 and is valid for predicting economic behaviour (Evans & Revelle, 2008). The PTS predict trusting behaviour in the Investment Game (Berg et al., 1995) more accurately than the general Big Five personality measure and consists of 26 items (Evans &

Revelle, 2008). Respondents indicate their level of agreement with statements about trusting behaviour on a Likert-scale with responses ranging from (1) “strongly inaccurate” to (6) “strongly accurate”.

Summary of subjective measures

A general stance to the subjective measurements of interpersonal trust pay attention to individual’s evaluation and expectation of another by the use of the Philosophies of Human Nature Scale (PHN: Wrightsman, (1964) and the social trust subscale of the European Social Survey (ESS: Breyer (2015). The research of Rotter provides an widely accepted definition of interpersonal trust: “An expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon (Beierlein, Kemper, & Rammstedt, Beatrice, 2012; Chun & Campbell, 1974; Rotter, 1967).” The ITS was the first instrument measuring generalized interpersonal trust. Rotter hypothesized that the Interpersonal Trust Scale (ITS) is most accurate for measuring trust in a highly uncertain, novel, or unstructured situations, where one's generalized expectancy is all one can rely on (Rotter, 1967). The work of Rotter is substantial to the trust literature and the development of interpersonal trust instruments. More recently it was shortened (KUSIV3). General trust was continued to be measured on an international level with the General Trust Scale and allowed for cross-cultural behavioural predictions with a focus on attitudes towards the general population. The single item world value scale is an alternative to the general trust scale. The complexity and multi-facades of interpersonal trust required a more focused approach, which lead to the development of the specific interpersonal trust scale and the propensity to trust scale allowing for individual differences in interpersonal trust scores to be observed (Evans & Revelle, 2008; Johnson-George & Swap, 1982).

Overview of objective trust measurements

Specific experimental tasks such as the trust game and its variants were proposed in the field of social psychology and behavioural economics as behavioural indications of trust during social interaction (Camerer & Weigelt, 1988). These measures have also been used to validate the previously mentioned subjective measures of interpersonal trust.

Table 14: Summary of objective trust measurements in experimental studies

Author	Experimental tasks	Task focus	Description
Camerer & Weigelt, (1988)	Lending game	studies reputation formation in an incomplete-information setting where a borrower (whose type is unknown) interacts with several lenders, but each bilateral interaction displays the elements of a trust situation. studies reputation formation in an incomplete-information setting where a borrower (whose type is unknown) interacts with several lenders, but each bilateral interaction displays the elements of a trust situation. This task was originally developed to test whether players behave honestly or dishonestly when incomplete information about the interaction partner was given. The effect of reputation formation on honest and dishonest behaviour was investigated with this task.	t. Subjects play an abstracted lending game: a B player lends or does not lend; then if Blends, an E player can pay back or renege. The game is played 8 times, and there is a small controlled probability that the E player's induced preferences make him prefer to pay back (but usually he prefers to renege). In sequential equilibrium, even E players who prefer to renege should pay back in early periods of the game, and renege with increasing frequency in later periods, to establish reputations for preferring to pay back. After many repetitions of the 8-period game, the actual play is roughly like the sequential equilibrium, except that E players payback later in the game and more often than they should. This behaviour is rational if B players have a "homemade" prior probability of .17 (in addition to the controlled probability) that E players will prefer to payback Each subject participated in one experiment. Subjects were paid the sum of the payoffs from all their decisions, in cash, at the end of the experiment. (Average payoffs were about $18 for the 2-1/2-hour experiment.) Each experiment had 75-100 repetitions of 8-period sequences, to give the subjects sufficient experience. Thus, the subjects observed 600-800 periods of play. Subjects play an abstracted lending game with two experimenters and 8 bankers. A mnemonic story that makes the discussion easier to follow is that an entrepreneur (E) (the subject) borrows from a different banker (B) each period for several periods. The entrepreneur (E) can have two types: an honest type E (called a "Y-type") and a dishonest type E (called “X-type”). In Camerer & Weigelt, (1988) experiment E’s type was randomly assigned to them, by an experimenter who chose a ball from a bingo cage with the type of E written on it. The type was announced to E (but not to the Bs) and was fixed for the entire 8 period sequence. However, both sides knew the prior probability that an Y or X-type would result. (After the sequence the type was then announced to the Bs). The experiment starts with three E subjects in one room with one experimenter and the eight B subjects (bankers) in another room with a second experimenter. The experimental instructions were read aloud to all the subjects simultaneously, everything about the experiment's structure was made transparent (except for E's type). Player B either lends an artificial currency (equal to $.01 for B subjects, and $.0015 for E subjects) or does not lend money; to player E, then if B lends, then player E can pay back or revoke payment. In each period, B decides whether to lend to E, at a fixed interest rate. If E gets a loan, he must decide whether to pay it back or not. Note how E's payoffs depended upon his or her type: If B chose to make a loan, a Y-type E got a higher payoff from paying back than from revoking. An X-type E got a higher payoff from revoking than from paying back. (Thus, the monetary payments subjects earned in the experiment induced nonmonetary preferences like those associated with honesty (Y-type) and dishonesty (X-type) in the natural world.)
Berg et al., (1995)	Trust game-binary choice	The investment game as an experimental task designed to investigate two questions: 1) is trust nascent to economic models of	A first agent, called the trustor, is given a monetary endowment X, and can choose which fraction p of it (zero being an option) will be sent to the second agent, called the trustee. The transfer p·X is then gone, and there is nothing the trustor can do to ensure a return of any kind. Before the transfer arrives into the trustee’s hands, the transfer is magnified by a factor K > 1 (e.g., doubled or tripled), both parties are not informed about this. That is, the trustor might send, say, $5 but the trustee receives $10 or more (always receives more than what was given). The trustee is free to keep the whole amount without repercussion. Crucially, however, the

		behaviour? 2) what factors increase (or decrease) trust during economic transactions?	trustee has the option to send a fraction q of the received transfer back to the trustor, hence honoring the trustor's initial sacrifice. Since p and q can in principle be any proportion, this is an infinite game, although in practice experimental implementations discretize the decisions, for instance requiring transfers to be integers. In the laboratory, roles are assigned randomly, the trustor-trustee matching is equally random, and interactions are computerized, one-shot, and anonymous.
Fehr et al., (1993)	Gift Exchange Game	Test the impact of fairness	Subjects either take on the role of the employer or the employee and play the gift exchange game with each other. Employers make wage offers which employees can repay with appropriate effort levels. Employees have no incentive to provide any effort above the minimum level, which, if anticipated by the employers, leads to minimum wages. However, both employer and employee are better off if the employer trusts the employee by offering a wage above the minimum and the employee pays back that trust by exerting a higher effort.
Lyons & Mehta, (1997)	Trading game	The role of trust in facilitating efficient exchange relations when agents are vulnerable to opportunistic behaviour.	Subjects either take on the role of the supplier or the seller and play the trading game with each other. After a previous, non-binding agreement, a supplier decides how much to invest (say, effort or capital) and then a buyer decides whether to pay as agreed or delay (unilaterally renegotiate the terms down). Other prominent examples have embedded trust-based interactions in more complex paradigms.
Glaeser et al., (2000)	Envelop drop	treeeeeeeee Evaluation of trust towards strangers	Subjects report valuations for a series of "envelope drops". Subjects are told that an envelope, addressed to the subject and containing $10, will be intentionally dropped by an experimenter's assistant. Envelopes will be dropped in different public places and under different conditions (e.g. sealed and stamped). For each place and condition of the envelope drop, the subject reports a valuation. Subjects have an option to either place a high or low value on the dropped envelopes. If a subject typically places a high value on the dropped envelopes, then the researchers infer that the subject is more likely to trust the strangers who will find the lost envelope. Glaeser et al, (2000) believe that the subject's valuation of such an envelope drop primarily measures confidence that a stranger in that location will return the envelope to the subject (by putting it into a mailbox). The procedure may also measure the subject's trust that the experimenter will carry out the envelope drop in the first place.
Bolton et al., (2004)	The Trust Dilemma	A trus A variant of the trust game specifically focusing on the relationship between buyers and sellers in the online trading market. The came captures all the critical features of online trade.	The subjects either take on the role of the supplier or the seller and play the Trust Dilemma game with each other. In this trust game, both the seller and the buyer are endowed with 35, if no trade takes place. The sellers offers an item for sale for 35, which has a value of 50 to the buyer. The seller's cost of providing the buyer with the item cost – the costs associated with executing the trade, shipping, handling, as well as production cost- is 20. So each successfully completed trade, creates a customer surplus of 15 and a net profit of 15 for the seller. If the buyer chooses to buy for the seller, the buyer send 35 to the seller, who then decides whether they will ship the item to the buyer or not. If he does not ship the item, then he receives the price plus his endowment of 35 as well as the surplus for a total of 70. If he ships he receives the price minus the costs, plus his endowment for a total of 50. If the buyer chooses not to buy, then no trade occurs.
Hale et al., (2018)	Experimental maze task	ff A behavioural task in virtual reality investigating specific trust	A behavioural task where subjects need to navigate through a virtual maze by making a series of decision on how to proceed. Before each decision the subject has the option to ask for advice from two virtual agents their briefly interviewed earlier. The researcher manipulated the virtual agents trustworthiness by verbal answers, non-verbal and vocal behaviour. The researchers measured how often subjects approach and follow the advice from each virtual agent. The main outcome of the experimental virtual maze task is that subjects follow the advice from the trustworthy virtual agent more than the untrustworthy virtual agent
Juvina et al., (2019)	Multi-arm trust game	An experimental paradigm investigating the affect multiple trustees on trust	The MATG is a game of strategic interaction combining features of two different games, the multi-arm bandit game (Robbins, 1952) and the trust game (Berg, 1995). The MATG is played between 4 players who interact repeatedly. One of the four players is randomly assigned the role of the Sender while the other three players are assigned the role of the Receiver. Over a series of rounds in the MATG, each player makes a set of decisions depending on their role in the game. At the start of each round both the Sender and Receiver each make an initial decision. First, the Sender is given a per-round endowment of 40 points. The Sender is then allowed to freely allocate their 40 point endowment between themselves and the Receivers. The Sender can give as much or as little of the 40 points as they wish to either themselves or to any of the 3 Receivers. As the Sender allocates their per-round endowment, each Receiver must decide to interact or not to interact with the Sender. If a Receiver decides not to interact with the Sender, then the

			Receiver will earn a random number of points selected from a distribution that is unknown to the Receiver. If a Receiver decides to interact with the Sender, then the Receiver will be given the number of points allocated to them by the Sender multiplied by 4. For example, if a Receiver decides to interact with a Sender and the Sender allocated 4 points to that Receiver, then the Receiver would be given 16 points. Additionally, Receivers who choose to interact with the Sender are allowed to return any number of their received multiplied allocation to the Sender. After all the Receivers have made their respective choices, the Sender is then notified of the choices made by each of the Receivers for that round. If the Sender allocated points to a Receiver who chooses not to interact with the Sender during that round, then the Sender is notified that they could not send their points to the Receiver during this round and the Sender is given back the points allocated to the Receiver. If a Sender allocated points to a Receiver who chooses to interact with the Sender, then the Sender is notified about the number of points allocated to the Receiver, the multiplied number of points that the Receiver was given, and how many points the Receiver returned to the Sender. The Sender is also told the total number of points earned during a given round. After the Sender observes the information about the Receivers the next round begins and the same procedure is repeated.

A review of experimental tasks in laboratory settings to measure interpersonal trust

The trust game (also known as the investors game) (Berg et al., 1995; Kreps, 1990), was established in the field of behavioural economics and is a popular technique for studying trust decision making (e.g., Alós-Ferrer & Farolfi, 2019; Glaeser et al., 2000; Jasielska et al., 2019). In socioeconomics research, for example an experimental study on "trust among internet traders", suggest that online markets are very "trust-unfriendly environments", yet many online trading platforms prosper. Bolton et al (2004), suggest that the trust game accurately describe the market institutions, however the consumers does not behave as selfish as economists typically assume and that traders care about morals and social impact. The use of one-shot games such as the trust game and its variants, in socioeconomic research enables psychologists, sociologists and experimental economists to identifying non-pecuniary motives that may drive economic behaviour (Bolton et al., 2004).

Trust games demonstrate the contrast between theoretical predictions of behaviour and the way that play occurs in practice. developed the *"Trust Game"* not only to measure trust but also many other related constructs such as fairness or reciprocity (Alós-Ferrer & Farolfi, 2019; Berg et al., 1995). Trust games measure the construct of trust in a social context in which the person makes decisions about the allocation of money that is distributed between himself and a stranger (Berg et al., 1995). In experimental studies from the field of behavioural economics there are many approved variants of the trust game for example, the lending game, binary-choice trust game, gift exchange game, trading game, envelope drop game etc. (Berg et al., 1995; Camerer & Weigelt, 1988; Fehr, 2009; Glaeser et al., 2000; Hale et al., 2018; Kreps, 1990b; Lyons & Mehta, 1997). These trust games are often used as experimental tasks to measure interpersonal trust behaviour which includes evaluations of experimental stimuli and performance of tasks and is often used to validate subjective trust measures (Berg et al., 1995; Cardoso et al., 2017; Jasielska et al., 2019). Table 14 provides a comprehensive overview of the most popular and most recent variations of trust games used to study interpersonal trust behaviour in experimental conditions. The Trust game is an established behavioural measure of interpersonal trust. Observable interpersonal trust behaviour during experimental trust games where a social interaction partner is present is the amount a participant is willing to send back during the trust game (e.g investment game), and also includes, for example, eye contact between social interaction partners, head movement, self-disclosure (Li et al., 2016; Rogers et al., 2015; Sluganovic et al., 2016). An experimental study investigating the relationship between mimicry and trust in virtual conversations agrees with previous research that mimicry can increase rapport, which lead to increased trust, hence the researchers suggest a VR mimicry paradigm to measure trust (Hale & Hamilton, 2016). Further research indicates that head motion, interpersonal distance between social interaction partners, advice-seeking behaviour, and a willingness to take a risk (Hale et al., 2018; Li et al., 2016; Peña & Yoo, 2014; Rosenberger et al., 2020) can be an indicator of trust. The experimental maze task (Hale et al., 2018), did not have a social interaction partner physically present, nor does research on chatbots. However the participants advice-seeking behaviour during uncertain circumstance and willingness to share information /self-disclosure in exchange for advice/guidance can serve as an indicator of interpersonal trust and correlate with a risk attitude (Derlega & Chaikin, 1977; Khawaji et al., 2015; Lee et al., 2018; Maloney et al., 2020). A conceptual model on privacy management suggest that the more personal information a participant chooses to disclose, the greater the level of interpersonal trust (Juvina et al., 2019). Juvina et al. (2019) contributes a novel prediction that trust will decrease when a trustor attempts to interact with a trustee but could not do so. The Multi-Arm trust game display differences between human-human trust and human-machine trust. We recommend that the subjective measurements of trust (trust questionnaires in Table 13) and objective measures of trust (experimental tasks_ Table 14) should be used in conjunction when studying interpersonal trust in an experimental setting.

Two research groups recent contributions stood out in the narrative literature review. The first group is Glaeser et al., (2000), they made use of three measurements, firstly a subjective measure of trust. Secondly, an objective behavioural measure with two experimental trust tasks (the trust game and the envelope drop).

Summary of trust measures

This review included subjective, objective, and behavioural measures of interpersonal trust in experimental research. The subjective measures (Table 13) include standardized valid and reliable paper and pencil questionnaires and self-reports. The most prominent subjective measure of interpersonal trust is the “Interpersonal Trust scale” (ITS)(Hale et al., 2018). In interpersonal trust research the *Trust Game* and its variations (Table 14) (Berg et al., 1995: were established in the field of behavioural economics. It is a popular technique for studying trust decisions and serves as an observational behavioural measure of interpersonal trust, with a monetary amount send back to the social interaction partner as an indication of interpersonal trust.

Appendix B

Study 1: Familiarity Manipulation Scripts

Guten Tag! Ich freue mich darauf, Sie beim Rundgang durch die Virtual-Reality-Kunstgalerie zu treffen. Meine Kunst wurde digitalisiert und ein Avatar von mir erstellt, damit wir uns in einem virtuellen Museum treffen können. Aber bevor wir uns treffen, möchte ich Ihnen ein wenig über mich erzählen. Ich bin Hannah Müller /Julia Köning, eine Künstlerin der modernen Kunst aus Stuttgart. Ich habe mit 7 Jahren zusammen mit meiner Schwester mit dem Kunstunterricht begonnen und war mein ganzes Leben lang gerne kreativ. Ich habe Kunst an der Universität studiert und erhielt 2020 eine Auszeichnung als kreativste Künstlerin. Mein Haus ist voll mit moderner Kunst aus der Region, geschaffen von Künstlern aus ganz Deutschland. Leider mag mein Partner Kunst nicht so sehr wie ich, aber zum Glück unterstützen sie mich trotzdem. Ich fahre jeden Monat nach Würzburg, um Zeit mit meiner Mutter zu verbringen und lokale Kunstgalerien zu besuchen. Ich hoffe, dass Ihnen meine Kunstausstellung gefällt. Bis bald!

Appendix D

Study 4: Emotional Manipulation – AEMT

A_ Autobiographical Emotional Memory Task (AEMT) – Negative condition

Offenlegung: Sie werden diese Seite mit nach Hause nehmen. Der Experimentator wird nicht lesen, was Sie geschrieben haben, und es wird nicht für die Datenanalyse verwendet.

Anweisungen:
Bitte denken Sie an eine negative Situation, in der Sie sich am unkontrolliertesten, ängstlichsten und/oder frustriertesten gefühlt haben. Das kann etwas sein, das Sie gegenwärtig erleben, oder etwas aus der Vergangenheit. Beginnen Sie damit, eine möglichst detaillierte Beschreibung des negativen Ereignisses aufzuschreiben, an das Sie sich erinnern können. Wenn Sie können, schreiben Sie Ihre Beschreibung bitte so, dass jemand, der dies liest, sich vielleicht sogar außer Kontrolle, ängstlich und/oder frustriert fühlt, nur weil er von der Situation erfährt. WELCHE negative Sache ist passiert, durch die Sie sich am meisten außer Kontrolle, ängstlich und/oder frustriert fühlen? WARUM fühlen Sie sich so unkontrolliert, ängstlich und/oder frustriert? Bitte schreiben Sie einige Minuten lang.

Danke, Sie können jetzt den Experimentator anrufen!

B_ Autobiographical Emotional Memory Task (AEMT) - Positive condition

Offenlegung: Sie werden diese Seite mit nach Hause nehmen. Der Experimentator wird nicht lesen, was Sie geschrieben haben, und es wird nicht für die Datenanalyse verwendet.

Anweisungen:

Anweisungen:

Bitte denken Sie an eine positive Situation, in der Sie sich besonders freudig und/oder glücklich gefühlt haben. Das kann etwas sein, das Sie gerade erleben, oder etwas aus der Vergangenheit. Beginnen Sie damit, eine möglichst detaillierte Beschreibung des positiven Ereignisses aufzuschreiben, an das Sie sich erinnern. Wenn Sie können, schreiben Sie Ihre Beschreibung bitte so, dass jemand, der dies liest, vielleicht sogar Freude und/oder Glück empfindet, wenn er nur von der Situation erfährt. WELCHE positive Sache ist passiert, die Sie am meisten erfreut und/oder glücklich gemacht hat? WARUM fühlen Sie sich so freudig und/oder glücklich? Schreiben Sie bitte mehrere Minuten lang.

Danke, Sie können jetzt den Experimentator anrufen!

Appendix E

Study 4: Specifications and SAM ratings of the positive IAPS and IADS-E

NEGATIVE CONDITION (NEA)												
Room	**Picture - Discription**	**Picture - code**	**Valence (mean)**	**Arousal (mean)**	**Dominance (mean)**	**Sound - Discription**	**Sound - Code**	**Valence (mean)**	**Arousal (mean)**	**Dominance (mean)**	**Sound time**	**Condition**
1	StarvingChild	1_9040	1.67	5.82	3.1	Suspense Thriller14	1_1367_2	4.64	4.27	1.68	0.06	1-Negative
2	Dog	2_9570	1.68	6.14	3.37	Suspense Thriller16	2_1380_2	2.56	5.16	2.10	0.06	2-Negative
3	OpenGrave	3_3005.1	1.63	6.2	2.77	Mystery BGM12	3_1362_2	3.50	6.17	1.82	0.06	3-Negative
4	Leopard	4_1310	4.6	6	4.37	BGM1	4_0008_2	5.44	5.36	1.96	0.06	4-Negative
5	Toddler	5_2095	1.97	5.25	3.7	Scared	5_0355_2	1.82	6.50	2.77	0.06	5-Negative
6	SadChildren	6_2703	1.91	5.78	3.15	Suspense Thriller3	6_0007_2	3.16	5.44	2.42	0.06	6-Negative
7	Soldier	7_9410	1.51	7.07	2.81	Piano9	7_1381_2	3.08	4.12	1.95	0.06	7-Negative
8	Woman	8_2375.1	2.2	4.66	3.75	Piano10	8_1381_b_2	3.21	4.54	1.95	0.06	8-Negative
9	AngryFace	9_2120	3.65	4.93	5.3	Thriller Suspense Film1	9_0209_2	2.29	5.83	2.73	0.06	9-Negative
10	DeadBody	10_3120	1.56	6.84	3.32	BabyCry3	10_1390_2	3.09	6.45	2.17	0.06	10-Negative
11	Riot	11_2691	3.9	5.65	5.54	BGM2	11_0010_2	5.63	4.83	1.94	0.06	11-Negative
12	War	12_2683	3.32	5.99	4.01	BGM18	12_1157_2	5.86	2.95	2.03	0.06	12-Negative
	Mean		**2.47**	**5.86**	**3.77**	**Mean**		**3.69**	**5.14**	**2.13**		
	Standard deviation		**1.09**	**0.70**	**0.91**	**Standard deviation**		**1.36**	**1.06**	**0.34**		
	high valence to low valence					low valence to high valence						
	Pictures from IAPS					Sounds from IADS-E						

Appendix F

Study 4: Specifications and SAM ratings of the negative IAPS and IADS-E

POSITIVE CONDITION (PEA)												
Room	**Picture - Discription**	**Picture - code**	**Valence (mean)**	**Arousal (mean)**	**Dominance (mean)**	**Sound - Discription**	**Sound - Code**	**Valence (mean)**	**Arousal (mean)**	**Dominance (mean)**	**Sound time**	**Condition**
1	Kitten	1_1460	8.21	4.31	6	Country Music4	1_1078_2	7.00	8.21	5.29	0.06	1 - Positive
2	Puppies	2_1710	8.34	5.41	6.55	Country Music1	2_1076_2	7.63	7.92	5.83	0.06	2 - Positive
3	Baby	3_2040	8.17	4.64	7.14	Country Music7	3_1085_2	8.09	7.77	6.50	0.06	3 - Positive
4	Seal	4_1440	8.19	4.61	6.05	Jazz1	4_1073_2	7.95	6.77	6.27	0.06	4 - Positive
5	Family	5_2340	8.03	4.9	6.18	Martial Music1	5_1114_2	7.75	7.92	6.21	0.06	5 - Positive
6	NeutBaby	6_2260	8.06	4.26	7.47	Pops9	6_1081_b_2	6.63	7.96	4.83	0.06	6 - Positive
7	Baby	7_2050	8.2	4.57	7.71	Country Music3	7_1077_b_2	7.67	6.38	5.38	0.06	7 - Positive
8	Babies	8_2080	8.09	4.7	7.08	Pops7	8_1080_2	7.86	6.59	5.73	0.06	8 - Positive
9	Bunnies	9_1750	8.28	4.1	6.15	Birds2	9_0419_2	7.86	4.55	6.68	0.06	9 - Positive
10	Baby	10_2070	8.17	4.51	7.33	Pops6	10_1079_b_2	7.17	7.63	5.08	0.06	10 - Positive
11	Beach	11_5833	8.22	5.71	6.97	Symphony2	11_1109_2	7.80	7.68	5.60	0.06	11 - Positive
12	Skier	12_8190	8.1	6.28	6.14	Pops10	12_1082_2	7.36	7.77	5.91	0.06	12 - Positive
	Mean		**8.17**	**4.83**	**6.73**	**Mean**		**7.56**	**7.26**	**5.78**		
	Standard deviation		**0.09**	**0.65**	**0.62**	**Standard deviation**		**0.44**	**1.04**	**0.57**		
	high valence to low valence					low valence to high valence						
	Pictures from IAPS					Sounds from IADS-E						

Appendix G

Study 4: Velten Statements Debriefing 1 (after NEA)

Instructions

Lesen Sie jede der folgenden Aussagen durch. Konzentrieren Sie sich beim Lesen der Aussagen, immer auf die jeweilige Aussage, welche Sie in diesem Moment lesen. Sie sollten nicht zu viel Zeit mit dem Lesen einer Aussage verbringen. Ihr Erfolg beim Erleben dieser Stimmung wird weitgehend von Ihrer Bereitschaft abhängen, die Idee in jeder Aussage zu akzeptieren und darauf zu reagieren und jeder Aussage zu erlauben, auf Sie einzuwirken. Versuchen Sie, auf das Gefühl zu reagieren, das durch jede Aussage suggeriert wird. Versuchen Sie sich dann vorzustellen, dass Sie definitv in **diesem Zustand/ in dieser Stimmung** sind oder **in diesen Zustand/ in diese Stimmung** kommen. Wenn es sich für Sie natürlich anfühlt, versuchen Sie, sich eine Situation vorzustellen, in der Sie ein solches Gefühl hatten.

Positive Velten Statements

1. Der heutige Tag ist weder ein besserer noch ein schlechterer Tag als jeder andere.
2. Allerdings fühle ich mich heute sehr gut.
3. Ich fühle mich leicht und unbeschwert.
4. Es könnte sich herausstellen, dass dieser Tag einer meiner guten Tage sein wird.
5. Wenn deine Einstellung gut ist, dann sind die Dinge gut und meine Einstellung ist gut.
6. Ich fühle mit fröhlich und lebendig.
7. Ich habe sicherlich Energie und Selbstvertrauen, welche ich teilen kann.
8. Im Großen und Ganzen habe ich wenig Schwierigkeiten klar zu denken.
9. Meine Freunde und meine Familie sind die meiste Zeit ziemlich stolz auf mich.
10. Ich bin in einer guten Stellung, um Dinge zum Erfolg zu führen.
11. Für den restlichen Tag werden Dinge sicherlich sehr gut verlaufen.
12. Ich bin froh, dass die meisten Menschen so freundlich zu mir sind.
13. Meine Urteile über die meisten Dinge sind gut überdacht.
14. Je mehr ich mich mit Dingen beschäftige, desto leichter werden diese für mich.
15. Ich bin voller Energie und Tatendrang – Ich fühle mich so als könnte ich eine lange Zeit ohne Schlaf auskommen.
16. Dies ist einer der Tage, an denen ich, fast ohne jegliche Anstrengung, Dinge erledigen kann.

17. Heute ist mein Urteilsvermögen scharf und präzise. Lassen Sie einfach jemanden versuchen mir etwas überzustülpen.
18. Wenn ich möchte, kann ich ganz extrem einfach Freunde finden.
19. Wenn ich mir etwas in den Kopf setze, kann ich Dinge ins Gute wenden.
20. Jetzt fühle ich mich enthusiastisch und selbstbewusst.
21. Es sollte Gelegenheit dafür geben, dass viele gute Zeiten kommen werden.
22. Meine Lieblingslieder gehen mir die ganze Zeit durch den Kopf.
23. Manche meiner Freunde sind so lebendig und optimistisch
24. Ich bin gesprächig - Ich fühle mich, als könnte ich mich mit fast jedem unterhalten.
25. Ich fühle mich voller Energie und fange an die Dinge, die ich tue, zu mögen.
26. Ich fühle mich als könnte ich laut loslachen – Ich wünschte jemand könnte mir einen Witz erzählen, um mir die Möglichkeit dazu zu geben
27. Ich fühle mich, bei allem, was ich tue, beschwingt.
28. Mein Gedächtnis ist heute in so guter Form wie selten.
29. Ich bin in der Lage Dinge akkurat und effizient zu erledigen.
30. Ich weiß sicher und gut, dass ich die Ziele, die ich mir gesetzt habe, erreichen kann.

Curriculum Vitae

Work experience

PhD Student — **University of Würzburg**

Biological Psychology, Clinical Psychology and Psychotherapy Department 08/2020 – 12/2023
Research on privacy protection in immersive virtual reality: evaluation and acceptance of social interaction partners' virtual reality, with Prof. Paul Pauli, Prof. Marc Erich Latoschick, Dr. Ivo Käthner.

Research Consultant — **Mthente Consulting**

Research Lead *01/2017 – 07/2023*
Meetings with clients, leading team meetings, formulating data collection plan, constructing and reviewing/approving data collection tools, analysing data, presenting data, writing reports. Projects usually involve governmental bodies, such as The South African Mining Authority (MQA) and the South African Tourism Industry (SAT), which include national and international stakeholders.

Program coordinator — **SACAP**

Psychology Department: Johannesburg, South Africa *03/2020 – 08/2020*
Coordination of professional programs (B.Psych Equivalent, BA Honours, B.Psych, Coaching, Higher Certificate). Lecturing and facilitating workshop implementations, handling student queries, amending student study plans, contributing input in module development, offering student counselling, attending disciplinary hearings. Member of college academic leadership team etc.

Post graduate lecture & research supervisor — **SACAP**

Psychology Department: Cape Town. *02/2019 -12/ 2020*
With a student-centered approach, preparing and delivering lectures (Research methodology, ethics, starting a practice, conflict management, preparation for the board exam etc.) to post graduate students. Supervising and grading post graduate honours research theses.

Post graduate lecture & research supervisor — **Varsity college**

Psychology Department: Cape Town *02/2019 -12/ 2020*
With a student-centered approach, preparing and delivering lectures (Research methodology, biological psychology, introduction to psychology) to post graduate students. Supervising and grading post graduate honours research theses.

Psychologist in Practice

Private practice: Cape Town, South Africa *01/2019 – 07/2023*
Health Professions Counsel of South Africa (HPCSA) registration number: PRC 0034568
Board of health funders (BHF) registration number: 0793256
Work includes individual therapy sessions (Anxiety, depression, trauma, behavioural problems, IMAGO couples counselling etc). Working with the schools in the Western Cape: Ned Doman High School, Kommetjie Primary, Sea Point Primary, Generations Schools (Hout Bay, Imhof, Bluemoon) Reddam Constantia. Western Cape Education Department (Central District) and Lotus River Community Clinic, Nedbank Head Office, The Student Hub etc.

Business Development Manager — **ActionCOACH**

ActionCOACH Business Coaching *02/2014 – 12/2016*
Sales Process development and implementation, marketing, managing client relationships (with CRM), building a database and tracking leads. Organizing small events, writing business alignment reports. Weekly training on marketing, sales, company culture, USP, personal growth of business. Attending network events etc.

Education

Dr.rer.nat. — **University of Würzburg**

Biological Psychology, Clinical Psychology and Psychotherapy Department 08/2020 - 07/2023
Marie Sklodowska Curie Fellowship. Research on privacy protection in immersive virtual reality: evaluation and acceptance of social interaction partners virtual reality, with Prof. Paul Pauli, Prof. Marc Erich Latoschick, Dr. Ivo Käthner.

M.Sc. — **University of the Free State**

Psychology department *01/2018 – 12/2019*

Research on the predictive factors of emotional intelligence among undergraduate university students, with Dr. Jacques Jordaan and Prof. Karel Esterhuize.

B.Psych.Equivalent **Cornerstone Institute**
Psychology department *02/2017 – 12/2018*
Research on suicide ideation.

Psychology Honours **UNISA**
Psychology department *01/2016 – 12/2017*
Research on depression

B.Soc.Sc **University of the Free State**
Psychology department *01/2011 – 12/2015*

Conferences and workshops

PRIMA workshop **University of Twente**
Research training and transferable skills *12/2020*

PRIMA Research event **NTNU, Norway**
Research training: biometrics and authentication 03/2021

PRIMA conference **KU Leuven, Belgium**
Presentation *05/2021*
State of the art of trust measurement

PRIMA workshop **KU Leuven, Belgium**
GDPR & DPIA *08/2021*

PRIMA conference **University of Madrid, Spain**
Presentation *03/2022*
Behavioural paradigm to measure trust in VR

IEEE VR **Shanghai Jiao Tong University**
Conference on Virtual Reality and 3D User Interfaces 2023
Presentation *03/2023*
Cronje, J., Lin, J., Kathner, I., Pauli, P., & Latoschik, M. E. (2023). Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm. IEEE Transactions on Visualization and Computer Graphics, 1–11. https://doi.org/10.1109/TVCG.2023.3247095

Internships

Triodos Bank **Zeist, Netherlands**
Research consultant *08/2021 – 12/2021*
My internship at Triodos Bank involved collaboration with the legal, IT, User Experience and security department. During my stay I gained an overview of the mobile on-boarding process from different departments and perspectives. In my role as researcher I investigated challenges regarding customer conversion rate during the mobile on-boarding process with a specific focus on customer biometric authentication attitudes and preferences. Online customer surveys were conducted, analysis of data was carried out, and results were reported, with recommendations to several departments.

NTNU **Gjøvik, Norway**
Visiting researcher *09/2022 – 12/2022*
My stay at NTNU in Gjøvik in Norway was hosted by the Department of Information Security and the Communication Technology (IIK) working group. I collaborated with another PhD student, designing an experiment to investigate the privacy paradox phenomenon. We plan to collect participant’s privacy preferences in different homecare environments under different conditions in virtual reality. The results will be utilised to develop privacy preference predictive models with the use of machine learning. I also gave an interactive seminar to the IIK working group on the "multidisciplinary nature of science". This seminar included the fundamentals of research in psychology, the use of virtual reality for

experimental research, and cyber psychology. We applied the theoretical learnings during the session, by designing several experiments (that use VR as a paradigm) to answer specific research questions in the field of health care.

Computer skills

Basic: R, Python, Unity3D, Latex
Intermediate: Jamovi, SPSS, LimeSurvey.
Expert: MS Office

Languages

Afrikaans: Mother tongue
English: Fluent
German: Basic A2
Dutch: Understand and read.

Interests

- Experimental design
- Virtual social interactions
- Virtual Humans
- Extended reality
- Cyber & Social security
- Human Factors and Human Behaviour
- Privacy and Security
- Data analysis & statistics
- Qualitative and Quantitative Research
- Psychology
- Cyberpsychology
- HCI
- User Research

Essential Skills

- Emotional Intelligent
- Leadership
- Creativity
- Human-centered thinking
- Curious
- Design Thinking
- Fast Learner
- Independent
- Team Player
- Empathetic
- Committed
- Structured
- Flexible

Publication List

Shared first author:

Cronje, J., Lin, J., Kathner, I., Pauli, P., & Latoschik, M. E. (2023). Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm. IEEE Transactions on Visualization and Computer Graphics, 1–11. https://doi.org/10.1109/TVCG.2023.3247095

Second author:

Lin, J., Cronje, J., Weinrich, C, Pauli, P., Latoschik, M.E. (2023) Visual Indicators Representing Avatars' Authenticity in Social Virtual Reality and Their Impacts on Perceived Trustworthiness. ISMAR TVCG

Manuscripts in preparation:

Cronje, J., Lin, J., Latoschik, M. E., Pauli, P., Käthner, I. (2023). The Impact of Cognitive Load on Trust Towards a Virtual Human in VR.

Cronje, J., Lin, J., Clements, M., Fahn., J, Käthener. I., Latoschik, M. E., Rodrigues. J. (2023). Construction and validation of a questionnaire measuring trust in specific avatars (SATS)

Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm

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

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 [60, 74], social interaction [62], education [18, 57], healthcare [25, 67], the workplace [41] 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 [44], 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 [62]. In many scenarios, such as online healthcare or financing, virtual humans can act as AI assistants, enhancing their social presence [39] and potentially increasing users' acceptance. The development of virtual humans, especially those with a realistic appearance, serves the purpose of enhancing virtual embodiment [61], recreating the real world [70], 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 xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

No lasting relationship can be established and maintained without trust [47,49], 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 [8,40]. Trust has been studied from many scientific perspectives, leading to a wide range of definitions [4, 10, 21, 27, 49]. 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 [48]. 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 [50]. We resonate with the definition of trust by Lee and See [38]: "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 [48], physical appearance (facial expressions, avatar behaviour, tone of voice) [42], 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 [77]. 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 [52]. 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 [40], 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.

[1]In this definition, an agent can be automation or another person that actively interacts with the environment on behalf of the person [38].

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 [63], and other alternatives such as the "General Trust Scale" (GTS) by Yamagishi and Yamagishi [76], the "KUSIV3" by Beierlein et al. [6], to name a few. In addition, other relevant scales, including the "Self-disclosure Index" (SDI) [46], the "European Social Survey" (ESS) [56], 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 [58], 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 [2,22,43], 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 [?] 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, 54, 59] 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 [39, 69] or combined self-reports with other measures [3, 26, 51]. 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, 51] 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 [51] 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, 53].

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, 61]. Given the often reported effects of virtual embodiments and presence on secondary factors such as emotional response [73] and especially trust [64], 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.

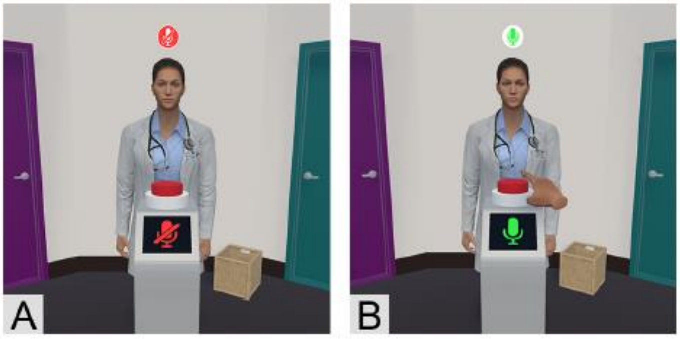


Fig. 2: A) Every time participants enter a new room, the virtual human will be muted. B) Participants can press the red button in front of the virtual human to unmute them if they ask for advice.

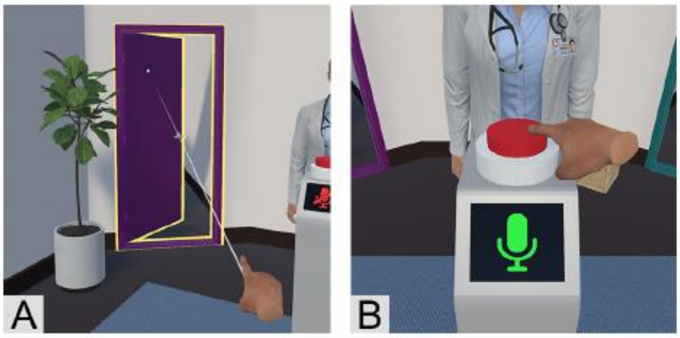


Fig. 3: A)Participants can point to a door and open it by pressing a button on their controllers. B)To seek advice from the virtual human, participants need to approach and unmute them.

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.

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, 55, 75]. 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

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

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	

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.

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 [66]. 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) [76] 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.

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 ($M = 5.17$s, $SD = 2.14$) and untrustworthy condition ($M = 5.43$s, $SD = 2.10$) 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.

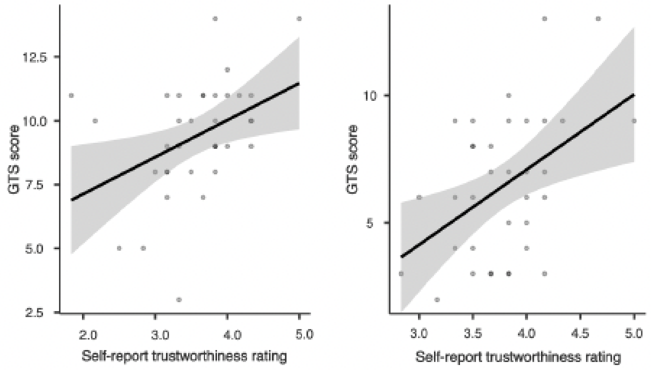


Fig. 5: Scatter plots for the correlation between the self-report trustworthiness rating and the general trust scale (GTS) score in the trustworthy condition (left) and the untrustworthy condition (right). The black lines in the plots represent the linear regression between the two variables.

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) [76].

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 [54,69] and vocal cues [71,72] 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 [76] 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 [2,43]. 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 [8,26,51], 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 [48] 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 [61], virtual body ownership [68], presence [66], congruency and plausibility [37], 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 [1,65], 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 [36,45]. 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] 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

[37] 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

[38] 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

[39] 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

[40] 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

[41] 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

[42] 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

[43] 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

[44] 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

[45] 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

[46] 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

[47] 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

[48] 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

[49] 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

[50] 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

[51] 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

[52] 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

[53] 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

[54] 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

[55] 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

[56] 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

[57] 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

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

[59] 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

[60] 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

[61] 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

[62] 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

[63] 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

[64] 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

[65] 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

[66] 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

[67] 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

[68] M. Slater. Towards a digital body: The virtual arm illusion. *Frontiers in*

Human Neuroscience, 2, 2008. doi: 10.3389/neuro.09.006.2008 8

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

[70] 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

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

[72] 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

[73] 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

[74] 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

[75] 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

[76] 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

[77] 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