IADIS International Journal on WWW/Internet
Vol. 22, No. 1, pp. 106-120
ISSN: 1645-7641

CONVERSATIONAL AGENTS AND THE IMPACT OF THEIR VISUALIZATION ON TRUST

Thomas Keller, Danny Wang and Elke Brucker-Kley
ZHAW Institute of Business Information Technology, Winterthur, Switzerland

ABSTRACT

This research delves into the significance of agents and trust within the metaverse, focusing on the dynamics of trust formation between individuals and agents in virtual environments, specifically in the realm of healthcare consultations. The primary objective is to pinpoint the factors influencing the trustworthiness of agents and explore their application in shaping digital avatars to bolster patient trust. The study draws on established psychological theories—including similarity attraction theory, attractiveness theory, exchange theory, attribution theory, and expectancy theory—to elucidate the intricate process of trust formation. Employing a laboratory experiment involving male participants aged 24-30, the study created diverse avatars with variations in visualization, voice, and gestures. The findings underscored that trust in agents is explicable through psychological theories, with participants exhibiting greater trust in avatars that demonstrated human-like features, competence, and expertise. Conversely, avatars featuring exaggerated or non-human characteristics were met with skepticism. Consequently, the study concludes that the application of psychological theories in the metaverse can effectively cultivate trust, emphasizing the need for avatar designs that incorporate elements enhancing trustworthiness. In essence, this research contributes valuable insights into the pivotal role of trust within the metaverse, particularly within healthcare contexts. The study advocates for the thoughtful consideration of trust-building factors in the design of digital avatars, providing practical recommendations to inform the creation of trustworthy virtual entities.

KEYWORDS

Agents, Virtual Reality, Trust

1. INTRODUCTION

The intricacies of trust formation between individuals are deeply rooted in diverse psychological theories and concepts (Simpson, 2007). Unraveling the complexities of this interplay necessitates a nuanced theoretical understanding of these psychological realms. When we extend this framework to the digital realm, where conversational agents play a prominent role,

a new layer of complexity emerges, introducing technological factors that further influence trust dynamics.

Recent studies, such as those by Seymour et al. (2020), underscore the pivotal role of realism in shaping trust towards agents. The degree of realism in an embodied bot significantly impacts human trust, thereby highlighting the intricate interplay between technology and trust dynamics. This nexus between technology and trust is particularly pertinent in the context of healthcare, as noted by Felnhofer et al. (2023), who identify a lack of trust in agents as a barrier, prompting patients to prefer physical visits to healthcare providers.

In light of the nascent construct of the metaverse, an intriguing question arises: can an agent within this virtual realm partially replace or supplement a real doctor? The uncertainty surrounding this question underscores the need to explore the applicability of psychological concepts related to interpersonal trust within the metaverse. This research embarks on such an exploration through a meticulously designed laboratory experiment, seeking to bridge the theoretical and practical dimensions of trust in the metaverse.

The laboratory experiment, a cornerstone of this work, aims to decipher the extent to which well-established psychological concepts, proven in interpersonal trust studies, can be transposed into the unique environment of the metaverse. By doing so, we endeavor to shed light on the fundamental question of how trust is cultivated between users and conversational agents in this digital space.

The outcomes of this research hold the promise of enriching our comprehension of the intricacies involved in designing digital avatars in healthcare settings to foster patient trust. Building on the foundations laid by prior literature, our study endeavors to provide actionable insights into the specific design elements crucial for engendering trust in the metaverse healthcare context.

As we delve into this rather uncharted territory, it becomes imperative to expand our theoretical framework. Recent works (Desideri et al., 2019; Jost et al., 2020; Vallverdú & Trovato, 2016) delve into the cognitive and emotional dimensions of human-robot interaction, offering insights into the factors influencing trust formation in technologically mediated relationships. Additionally, studies (Chérif & Lemoine, 2019; De Visser et al., 2016; Kulms & Kopp, 2019; Natarajan & Gombolay, 2020) on the impact of anthropomorphic features on user trust in virtual agents adds a valuable layer to our understanding of the role of visual design in shaping trust perceptions.

In synthesizing these diverse strands of literature, this research aims to not only contribute to the theoretical discourse surrounding trust in the metaverse but also to provide practical recommendations for healthcare in this digital frontier. By elucidating the intricate dance between psychological theories and technological factors, we aspire to guide the design and implementation of conversational agents in a manner that fosters trust and enhances user experience within the metaverse.

2. OBJECTIVE

The goal of this work is to investigate the relevance of influencing factors on the trustworthiness of an agent in the metaverse. For this purpose, a corresponding VR application is designed. Accordingly, the research question for this work is: "How should an agent be designed regarding

visualization, voice, and gesture for health consultations in order to foster trust among patients?".

In a first step, the relevant psychological theories are identified and assessed. Based on this, a VR app is developed in a second step, with which the relevance of the possible influencing factors is determined with a laboratory experiment. In a final step, statements are made and critically assessed based on the collected data.

3. RELATED WORK

Avatars and agents represent one of the core elements of the metaverse, enabling a self-presence for the user in the former case (Blockchain Research Lab, 2003) and an embodied bot in the latter case. This serves as the basis for social interaction in the metaverse. Visualizations can be anthropomorphic, abstract, animalistic as well as fantasy creatures. Non-verbal and paraverbal features can be designed to match the visualization (Sajjadi et al., 2018). On this basis, human-computer interaction is redefined.

Trust in an agent in the metaverse is believed to be influenced by various psychological theories and constructs. These may include factors such as attractiveness, degree of realism, or similarity to the user (Aljaroodi et al., 2019). According to Seymour et al. (2018), the term "realistic visual presence" (RVP) is representative of increased trust as well as willingness to interact with an agent.

Felnhofer et al. (2023) discuss a meta-analysis that examined the social responses of users towards virtual humans in immersive VR. The study aimed to determine whether the perceived agency of a virtual human, whether controlled by a person (avatar) or a computer (agent), affects social presence, evaluation of the virtual entity, and behaviours towards it. The analysis considered factors such as task type, appearance of the virtual entity, type of interaction, and agency manipulation. The results indicated that avatars were favoured over agents in terms of social presence. However, no significant effects were found for behavioural outcomes. Further analysis revealed that the differences in social presence were larger for neutral tasks compared to negative tasks, and differences in evaluation were more pronounced for positive tasks compared to mixed and neutral tasks. In summary, the findings suggest that deliberate social responses, such as social presence and evaluation, are influenced by the perceived agency of virtual humans. However, automatic behaviours do not appear to be affected. The study emphasizes the need for consistent conceptualizations of key variables in future research.

The individual perception of presence and immersion is highly relevant. The study by Doerner et al. (Doerner et al., 2022) shows that the subjective illusion of presence in a virtual environment is essential for the feeling of immersion. The degree of immersion therefore describes the extent to which a user feels involved in VR through stimulation of human perception. These are influenced by the perception of visualization, acoustics, haptics and movements, among other things. According to Sherman and Craig (2020), a distinction can also be made between mental and physical immersion. Mental immersion refers to the subjective feeling of presence in VR. Physical immersion, on the other hand, refers to the synthetic stimulation of the body sensation through the use of the technology. According to various studies, it is possible to measure the presence of test subjects in an experiment (Slater et al., 1994; Witmer & Singer, 1998). This can be measured by real stimulators of perception, such as visibly realistic movements, or by physiological factors, such as increased pulse (Doerner et al., 2022).

3.1 Avatars

Digital avatars are one of the core elements of the metaverse and allow users to represent themselves using a desired identity. This serves as the basis for social interaction in the metaverse. Avatars can be very diverse: self-presenting avatars, abstract avatars and non-human avatars are three examples of avatar types. Extensive configuration options such as the appearance or emotions allow the avatars to be customized according to individual preferences. On this basis, interactions between computers and humans can be designed in a new way (Kyrlitsias & Michael-Grigoriou, 2022; Seinfeld et al., 2021). In contrast to digital avatars, presence is also referred to as social presence according to Biocca & Chad (Biocca & Harms, 2002). This is defined by the feeling for other people or the degree of awareness of the presence of another person, being or type of intelligence. The acceptance of avatars is influenced by the context (Freeman & Maloney, 2021). For example, a photorealistic avatar in an arcade environment such as Sandbox will achieve a deeper presence and immersion of the user. It can thus be deduced that the reverse is true in a realistic setting such as a healthcare consultation. The major technological leaps in recent years now make it possible to create digital avatars that are almost photorealistic (Weidner et al., 2023). The entertainment industry in particular has been able to drive the creation of realistic faces in real time based on computer-generated technologies (Seymour et al., 2018).

3.2 e-Health

Telemedicine enables the interaction between doctor and patient, who are located in geographically different places (Sparwasser et al., 2022). The metaverse offers a way to improve existing telemedicine services (Skalidis et al., 2022). This includes interactions in the context of a health consultation. The visualization of the health advisor, whether human or bot, takes the form of an avatar (Rheu et al., 2020). This makes it possible to achieve a stronger persuasive effect than through conventional telemedicine tools, such as pure telephone consultation. VR goggles are also used in an attempt to provide a better patient experience in a three-dimensional environment (Bansal et al., 2022). In addition, the use of this technology can also improve access to healthcare in remote environments. People who suffer from physical limitations also typically benefit. Because the interaction is virtual via the metaverse, savings in travel costs and waiting time can also be realized. Virtual health consultation also reduces the risk of infectious disease transmission between patients and medical staff.

Another area of application in healthcare is virtual training and continuing education (Thomason, 2021). The benefits in terms of staff training are far-reaching. Through the use of VR, learners in the metaverse can gain a better overview of replicating actual procedures. This can illustrate processes in a more accessible way and enable interactive exchanges within lessons.

3.3 Psychological Theories

To model a person's trust in an agent, various social psychological concepts are used. These concepts deal with the way people interact with each other to build interpersonal relationships and trust. Based on this, it will be determined in relation to the research question how the digital avatars should be visually designed. The following sections cover five for this study considered theories.

3.3.1 Similarity Attraction Theory

At its core, similarity attraction theory posits that individuals are inherently inclined to be drawn to those who share similarities with them (Montoya & Horton, 2013). This foundational principle suggests that commonalities foster attraction, creating a sense of connection, while differences may introduce a level of discomfort. The notion that shared socio-demographic characteristics, such as race, gender, or education level, can amplify mutual attraction adds a nuanced layer to this theory (Roebken, 2010).

Roebken's insights underscore the broader spectrum of factors influencing attraction, highlighting the role of socio-demographic parallels in shaping interpersonal connections. This perspective broadens the scope of similarity beyond individual traits, recognizing the impact of societal constructs on the dynamics of attraction.

However, amidst the compelling arguments supporting the significance of similarities, it is essential to acknowledge the contested viewpoint that similarity in values and attitudes might not be the exclusive or primary factor in fostering attraction between individuals (Montoya & Horton, 2013; Santee, 1976). This divergence in opinion prompts a deeper exploration into the multifaceted nature of attraction, challenging the assumption that shared beliefs and values universally drive interpersonal connections.

Santee's skepticism invites scrutiny into the complex interplay of factors contributing to attraction, suggesting that while similarities in socio-demographic characteristics might create an initial affinity, the role of values and attitudes requires a more nuanced examination. This debate enriches the ongoing discourse on similarity attraction theory, emphasizing the need to consider a broader spectrum of elements that contribute to the intricate dance of human connections.

3.3.2 Attractiveness Theory

Samuel Frenning's attractiveness theory, as elucidated by Fink and Penton-Voak (2002), delves into the intricate process by which individuals assess the appeal of others and the consequential impact on interpersonal relationships. In contrast to the similarity-attraction theory, Frenning's framework posits a multi-faceted approach to attractiveness, incorporating three fundamental factors: physical, social, and task-related elements.

- **Physical Factors**: Frenning's theory recognizes the significant role of physical attributes in the evaluation of attractiveness. This extends beyond conventional notions of beauty and encompasses a broad spectrum of physical qualities that individuals consider when forming impressions of others.
- **Social Factors**: Beyond the physical realm, Frenning's theory acknowledges the importance of social dynamics in attractiveness evaluations. Social factors encapsulate qualities such as sociability, communication skills, and the ability to navigate interpersonal relationships effectively. These elements contribute to the overall appeal of an individual beyond their physical appearance.
- **Task-Related Factors**: Frenning's model expands the scope of attractiveness assessment to include task-related factors. This dimension underscores the significance of competence, skills, and capabilities in influencing how individuals perceive one another. Competence and proficiency in various tasks contribute to an individual's overall attractiveness.

In addition to these triadic elements, Frenning's attractiveness theory incorporates four guiding principles: closeness, similarity, reciprocity, and reinforcement. These principles act as crucial components shaping the dynamics of interpersonal relationships.

- **Closeness**: The proximity and emotional closeness between individuals are pivotal aspects influencing how they evaluate each other's attractiveness. Frenning recognizes that physical and emotional proximity contribute to the overall perception of appeal.
- **Similarity**: While differing from the similarity-attraction theory, Frenning's model acknowledges the role of similarity in certain contexts. Similarity in interests, values, or goals can enhance the perceived attractiveness of individuals to each other.
- **Reciprocity**: The give-and-take nature of relationships, captured by the principle of reciprocity, is highlighted in Frenning's framework. Mutual exchanges and responsiveness contribute to the positive evaluation of attractiveness.
- **Reinforcement**: Positive reinforcement of behavior and interactions reinforces mutual trust and contributes to the overall attractiveness of individuals within the social dynamic.

In essence, Frenning's attractiveness theory provides a comprehensive lens through which to understand the multifaceted nature of interpersonal evaluations. By incorporating physical, social, and task-related factors, along with guiding principles, the theory enriches our comprehension of how individuals navigate the complex landscape of attraction and build trust in their relationships.

3.3.3 Exchange Theory

Exchange theory, as articulated by Cook et al. (2013), offers a profound perspective on interpersonal relationships, shaping our understanding of social interactions through the lens of cost-benefit analysis. This theoretical framework posits that human interactions are essentially a series of exchanges wherein individuals weigh the costs and benefits to maximize positive outcomes.

In the contemporary landscape, exchange theory stands as one of the most pivotal theoretical views of social interaction and social structure. Its significance has been underscored by the contributions of sociologists such as George Homans (Homans, 1974), Peter Blau (Blau, 1968), and Richard Emerson (Emerson, 1964) during the transformative period of the 1960s to the 1970s.

For George Homans, a key figure in the development of exchange theory, the focus is twofold. First, he emphasizes interpersonal exchange, conceptualizing social interactions as a dynamic exchange of rewarding and costly activities between at least two individuals. This exchange forms the foundation of interpersonal relationships, where individuals engage in a reciprocal give-and-take to achieve positive outcomes. Second, Homans directs attention to the individual behavior of actors as they navigate these exchanges. This emphasis on individual agency highlights the active role each person plays in shaping and influencing the dynamics of social exchange.

In the framework of exchange theory, interpersonal relationships become dynamic arenas where individuals strategically engage in actions to maximize their own interests and minimize negative consequences. The essence of positive development, within this perspective, lies in the mutual benefit derived from the exchange process. When both parties engage in actions that are rewarding and mutually beneficial, the relationship flourishes, fostering positive outcomes for all involved.

As exchange theory continues to be a cornerstone in the study of social interactions, its enduring relevance lies in its ability to offer insights into the intricate dance of human relationships. By framing interactions as a calculated exchange of costs and benefits, the theory not only provides a theoretical lens through which to understand social dynamics but also lays

the groundwork for exploring the complexities of reciprocity, mutual gain, and the evolution of relationships in various social contexts.

3.3.4 Attribution Theory

Attribution theory, as expounded by Hewett et al. (2018), delves into the intricate processes by which individuals utilize information to construct causal relationships for various events. Within this framework, attribution encompasses the gathering and amalgamation of information to formulate judgments regarding causality. Notably, the emphasis lies on subjective explanations, prioritizing perceived correctness over objective accuracy, as highlighted by Raab et al. (2010).

In the context of this theoretical foundation, a key dimension gaining prominence is controllability, a concept underscored by Hatzakis (2009). The controllability dimension focuses on the perceived degree of influence the actor has over the outcome, thereby shaping subjective judgments and subsequent behaviors.

Central to the controllability dimension is the pivotal notion that individuals assess the level of control they believe they possess in a given situation. When an actor perceives low control over the outcome, a sense of helplessness and vulnerability ensues. This perception of limited influence over events can have profound implications, particularly in the realm of trust dynamics.

The intricate interplay between perceived control and trust is evident in situations where individuals feel their actions have minimal impact on the outcome. In such instances, the perceived helplessness and vulnerability stemming from a lack of control exert a negative influence on trust. This linkage between controllability and trust underscores the nuanced psychology involved in interpersonal relationships and decision-making processes.

Understanding the dynamics of controllability within attribution theory becomes especially pertinent in scenarios where trust is a critical factor. Whether in professional collaborations, personal relationships, or broader societal contexts, individuals' assessments of their ability to influence outcomes can significantly shape the fabric of trust.

In essence, the exploration of attribution theory, with a focus on controllability, provides a lens through which we can unravel the complex tapestry of human cognition and interpersonal dynamics. By acknowledging the subjective nature of causal judgments and their impact on perceived control, this framework contributes valuable insights into the psychological underpinnings of trust and vulnerability in diverse contexts.

3.3.5 Expectancy Theory

Expectancy theory, as articulated by Behling and Starke (1973), serves as a foundational framework rooted in the assumption that individuals tend to choose actions that align with their subjective expectations. This fundamental premise finds significant resonance in the realm of trust formation, as it implies that people are inclined to trust individuals who are perceived as most likely to fulfill their expectations.

A crucial dimension within expectancy theory is the prelude of subjective evaluations concerning the anticipated results and consequences of actions, as highlighted by Rank (1997). This introspective aspect adds layers of complexity to the trust-building process, emphasizing the cognitive assessment individuals undertake before placing their trust in someone.

The notion of utility maximization is intrinsic to expectancy theory and resonates with the idea of rational action. Many individuals, when navigating choices and forming expectations, engage in a form of utility maximization—a process driven by the desire to optimize personal

satisfaction. In the literature, this theory is interchangeably referred to as value-expectancy theory or utility theory, contingent on the contextual focus.

Utility theory, as expounded by Bühlmann et al. (1969), further extends the concept of utility maximization beyond individual actions to encompass product or service selection. In this context, individuals are postulated to choose products or services based on the expectation of attaining the maximum value for their investment.

Transposing these theoretical underpinnings into the domain of trust-building, a logical deduction emerges: individuals are inclined to trust service providers or counterparts who are perceived as most likely to fulfill their personal satisfaction. This linkage underscores the reciprocity between expectancy theory and trust dynamics, emphasizing the role of anticipated outcomes and the pursuit of maximum utility in shaping trust relationships.

In essence, expectancy theory offers a comprehensive lens through which to understand the intricacies of trust formation. By acknowledging the interplay between subjective expectations, utility maximization, and the rational evaluation of potential outcomes, this theoretical framework enriches our comprehension of how individuals navigate trust in a myriad of contexts, from interpersonal relationships to consumer choices.

4. CONCEPTION OF THE VR APP AND THE AGENT

The creation of the agent and its variations is based on the technological and psychological criteria. When creating the avatars, it was also ensured that they were created as analogously as possible to the homogeneous characteristics of the subjects. Accordingly, only male avatars were created in the context of this work, analogous to the selection of the subjects. In this way, the result with regard to the creation of trust through voice, appearance and interaction is not distorted by other psychological influencing factors, e.g. gender. Table 1 describes the characteristics of the avatars that were considered during the creation process, while Figure 1 shows their visual representation.

Table 1. Avatar visualization and animation features

#	Style	Type	Voice	Gestures	Lip-Sync
1	Realistic	Young male	Human voice	Complex animations	complex
2	Realistic	Old male	Human voice	Complex animations	complex
3	Cartoon	Young male	Human-like voice	Simple animations	simple
4	Cartoon	Old male	Human-like voice	Simple animations	simple
5	Non-Human	Animal	Generated voice	No animations	simple
6	Non-Human	Phantasy	Generated voice	No animations	simple

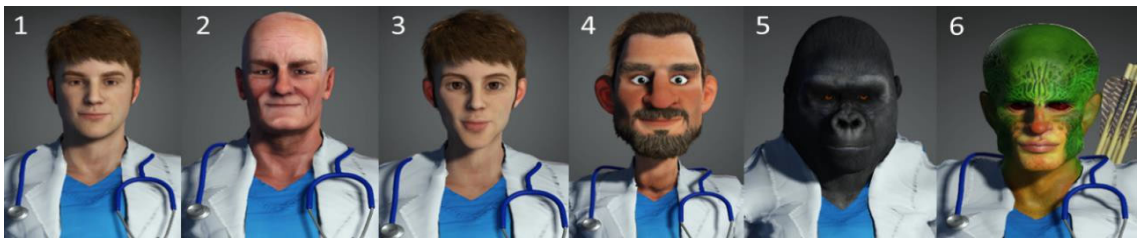

Figure 1. Avatar visualization

The interaction with the test persons in the context of a simulated consultation takes place in a practice room, which is characterized by elements such as medicine boxes or disinfection products (Figure 2).

Figure 2. Doctor's office for the simulated consultations

Our working hypothesis is based on the following assumptions:

- The similarity-attraction theory is to be taken into account by creating avatars that are supposed to be similar to the subjects. This can be done through appearance, voice or non-verbal movements.
- The attractiveness theory should be taken into account through the attractiveness of the avatars. This can also be achieved through an attractive appearance, voice and non-verbal movements.
- The exchange theory should be taken into account through the meaningful design of the interaction. An appropriate dialog should give the respondent the feeling that their needs have been met during the conversation.
- The attribution theory should be taken into account through the diversity of the avatar creation. In this way, different subjective experiences of the test subjects from the past are to be examined and considered for the subsequent interactions.
- The expectation theory is to be taken into account through the design of avatars in different age groups. The aim is to investigate whether the age of the avatar influences expectations and consequently trustworthiness.

5. THE LABORATORY EXPERIMENT

The selection of 20 subjects is limited to males with a university degree or university students between the ages of 24 and 30. After acceptance, they were informed about the content and goal of the experiment. This will give them a rough overview of the course and design of the experiment in advance. In addition, the test persons are informed that the experiment can be carried out decentrally at the respective place of residence of the test person. On the one hand, this aims to increase motivation and willingness to participate. On the other hand, it ensures that the experiment can be conducted in a controlled environment without disturbing factors.

Immediately before the experiment is conducted, the subject is prepared for the upcoming interactions with the agents. His role as a patient in a health consultation is explained and the explanation of the evaluation form shows which factors have to be considered and evaluated during the interactions. The subject goes through the six variants of the consultation in random order and fill out the corresponding questionnaires in each case as shown in Figure 3. The three free variables concern the appearance of the agent, his voice, his gestures. The environment, the practice room, remains constantly realistic. The dependent variable is the perceived trust. The questionnaire contains five additional questions about the reasoning of trust, which are based on the psychological models.

During the laboratory experiment, covert observation of the test person takes place. Individual statements and reactions of the test person during the experiment are recorded. Subsequently, the subject is given an evaluation form to assess his experience. In this way, the subject is to record his experience as well as individual thought processes according to predefined criteria.

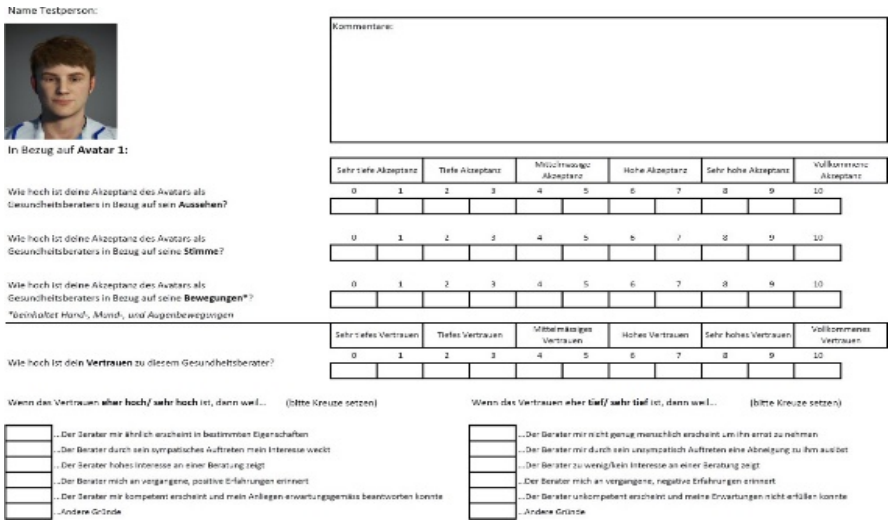

Figure 3. Questionnaire in its original German version

6. RESULTS

Figure 4 shows that the formation of trust towards the realistic avatars can be explained by different psychological theories. The high trustworthiness of Avatar 1 is mainly characterized by the expression of three psychological theories: The Similarity Attraction Theory, the Attraction Theory, and the Exchange Theory. The strong expression of the Similarity Attraction Theory can be justified by the fact that Avatar 1 was created as a young health care worker visually similar to the homogeneous group of subjects. Out of 20 subjects, 13 sensed many similarity features in Avatar 1 during the laboratory experiment, especially with regard to his appearance and age. His realistic appearance eliminated many irritating confounding factors that occurred during interactions with the other avatars. The attractiveness theory is even more

pronounced. 17 subjects found his appearance likeable and friendly. According to various statements, his human voice in particular generated a lot of sympathy among the subjects. Through his friendly appearance, he visibly showed a high level of interest in a health consultation, which significantly increased the subjects' willingness to interact. 14 subjects felt most noticed mainly because of his extroverted manner. As a result, beneficial interactions in the sense of the exchange theory were most pronounced with this avatar of all avatars. With an average score of 7.8 out of 10, Avatar 1's trustworthiness was rated as high by the subjects.

Avatar 2 has a slightly higher trustworthiness, which achieved an average score of 7.95 out of 10. The underlying psychological theories that characterize the trustworthiness of Avatar 2 are striking. The visibly higher age of Avatar 2 implied 15 subjects with a higher level of competence and more extensive expertise. For obtaining services concerning their own health, they are more likely to trust an experienced health advisor, as he or she has a more qualified impression. As a result, the expectation for expert health advice was highest for Avatar 2, which is why the expectancy theory was highest for him. 13 subjects further described their positive experiences with elderly service providers in different contexts. According to the subjects' subjective experiences, experts and specialists tend to be people of an older age, which is why they appear to be more trustworthy for obtaining a service. As a result, the attribution theory is also highest for Avatar 2. 11 subjects found his calm voice very pleasant and benevolent, which also confirms the attractiveness theory.

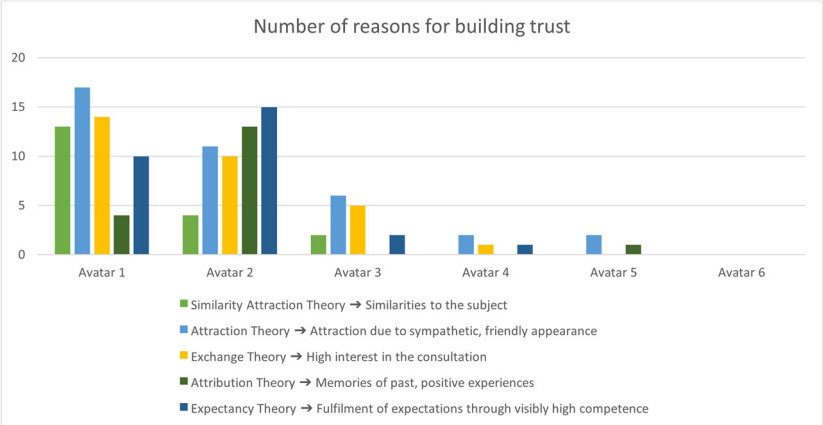

Figure 4. Reasons for building trust in the avatars

Avatar 3 elicited very mixed reactions from the subjects. Because his voice was pitched higher based on Avatar 1's voice, 6 subjects felt that interacting with him was friendly and inviting. His cartoon-like appearance caused some of the interaction with him to be described as humorous, resulting in a certain degree of attraction and trust building according to the attractiveness theory. However, the audibly artificial voice as well as the slightly exaggerated representations of facial features triggered a certain feeling of discomfort in some subjects. The reduced degree of realism caused 9 subjects to no longer perceive it as a lifelike human being. The slight discrepancy with the human appearance thus led to the formation of distrust in relation to the similarity-attraction theory. His high-pitched voice also caused 9 subjects to perceive him as an adolescent, thus questioning his competence as a health care provider. Therefore, following the expectancy theory, expectations toward him were low, which negatively affected his trustworthiness. With an average score of 5.05 out of 10, his trustworthiness was rated as mediocre.

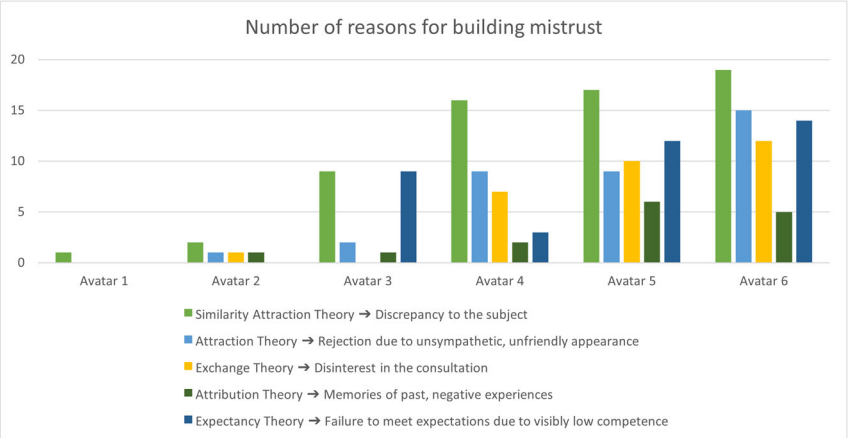

Figure 5. Reasons for mistrusting the avatars

Figure 5 shows how distrust increases with avatar 4. This is the first avatar for which trustworthiness was rated as low by the test subjects. The older cartoon avatar triggered low acceptance as a health advisor due to the exaggerated expressions of the facial features. Due to a deep voice pitch, his voice seemed inhuman and sinister to most subjects, which is why 16 subjects showed a strong rejection toward him according to the similarity-attraction theory. He thus appeared unattractive and unappealing for a trustworthy interaction in all areas, which led to the formation of distrust in 9 subjects according to attractiveness theory. The reduced level of non-verbal movements was interpreted by 7 subjects as a lack of interest in a health consultation, indicating a negative expression of the exchange theory. In general, his trustworthiness was rated low, with an average score of 3.7 out of 10. However, due to his advanced age, his professional competencies were not conspicuously questioned.

With an average value of 2.9, the trustworthiness of Avatar 5, which represents the first non-human avatar, was rated even lower. The gorilla triggered strong discomfort and rejection in 17 subjects due to its static movements and robotic voice. Thus, a strong negative expression according to the similarity-attraction theory is evident. This made it difficult for 10 subjects to perceive a serious interaction guidance in the context of a health consultation, which led to the formation of distrust according to the exchange theory. Based on diverse subjective experiences, the sight of the gorilla triggered a scary feeling in 6 subjects. Attributing to the attribution theory, this significantly reduced the trustworthiness of the gorilla towards these subjects. However, it should also be emphasized that the gorilla elicited sympathy in 2 subjects. The gorilla was received by them with humor, which is why the attractiveness theory has a small influence here. Beyond that, however, the question arose to what extent a gorilla would represent a competent and trustworthy health advisor in practice. According to the expectancy theory, 12 subjects were suspicious of whether a gorilla could trustworthily follow up on a patient's needs.

Avatar 6 has the lowest trustworthiness, which triggered distrust in all test subjects with an average value of 1.55 out of 10. The negative expression of the psychological theories took place to a high degree with this fantasy avatar. All reference to a human being is absent with this avatar. The unrealistic and inhuman representation seemed completely unsuitable for 19 test persons for a trustworthy execution of a health consultation. Due to the non-existent movements and the monotonous, robotic voice, no sympathy could be developed for him either, which is why 15 subjects found him unfriendly according to the attractiveness theory. For 14 subjects, the avatar seemed incompetent and unserious as a health advisor, which led to the formation of mistrust towards him according to the expectancy theory.

7. CONCLUSION

From the results of the laboratory experiment, it can be concluded that the psychological theories for building trust also find their applicability in the metaverse. The evaluation of the results show that the test persons trust or distrust the avatars in the context of a health consultation based on different psychological theories. It can be stated that the level of the individual theories varies depending on the avatar. In the case of a virtual health consultation, which is perceived as trustworthy and positive, a strong influence of the similarity-attraction theory and the expectancy theory can be observed. Avatars tend to be trusted more when they show basic similarities to human features. A visible deviation of avatars in this respect significantly reduces their trustworthiness. Expectation is also pronounced, especially when receiving a healthcare service. The avatars exhibit higher trustworthiness if they obviously give the impression that they have the assumed competencies and expertise. It should be emphasized, however, that the high expression of these theories is only confirmed for the group of subjects studied. It is conceivable that, for example, for a younger group of subjects, e.g., children, adolescents, high external discrepancies with a human being such as a cartoon avatar may trigger higher trust.

REFERENCES

Aljaroodi, H. M. et al., 2019. Avatars and Embodied Agents in Experimental Information Systems Research: A Systematic Review and Conceptual Framework. *Australasian Journal of Information Systems*, Vol. 23. https://doi.org/10.3127/ajis.v23i0.1841

Bansal, G. et al, 2022. Healthcare in Metaverse: A Survey on Current Metaverse Applications in Healthcare. *IEEE Access*, Vol. 10, pp. 119914-119946. https://doi.org/10.1109/ACCESS.2022.3219845

Behling, O. and Starke, F. A., 1973. The Postulates of Expectancy Theory. *Academy of Management Journal*, Vol. 16, No. 3, pp. 373-388. https://doi.org/10.5465/254999

Biocca, F. and Harms, C., 2002. *Defining and measuring social presence: Contribution to the Networked Minds Theory and Measure*.

Blau, P. M., 1968. Social exchange. *International encyclopedia of the social sciences*, Vol. 7, No. 4, pp. 452-457.

Blockchain Research Lab, 2023. Avatars: Shaping Digital Identity in the Metaverse (Scientific report). https://www.blockchainresearchlab.org/wp-content/uploads/2020/05/Avatars-Shaping-Digital-Identity-in-the-Metaverse-Report-March-2023-Blockchain-Research-Lab.pdf

Bühlmann, H., Loeffel, H. and Nievergelt, E., 1969. Nutzentheorie und Ihre Anwendung. In H. Bühlmann, H. Loeffel and E. Nievergelt (Eds.), *Einführung in die Theorie und Praxis der Entscheidung bei Unsicherheit: Unterlagen für einen Kurs der Schweizerischen Vereinigung für Operations Research* (pp. 29-43). Springer. https://doi.org/10.1007/978-3-642-80553-0_3

Chérif, E. and Lemoine, J.-F., 2019. Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche et Applications En Marketing (English Edition)*, Vol. 34, No. 1, pp. 28-47. https://doi.org/10.1177/2051570719829432

Cook, K. S. et al., 2013. Social Exchange Theory. In J. DeLamater and A. Ward (Eds.), *Handbook of Social Psychology* (pp. 61-88). Springer Netherlands. https://doi.org/10.1007/978-94-007-6772-0_3

De Visser, E. J. et al., 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, Vol. 22, No. 3, pp. 331-349.

Desideri, L. et al., 2019. Emotional processes in human-robot interaction during brief cognitive testing. *Computers in Human Behavior*, Vol. 90, pp. 331-342.

Doerner, R. et al., 2022. Introduction to Virtual and Augmented Reality. In R. Doerner, W. Broll, P. Grimm and B. Jung (Eds.), *Virtual and Augmented Reality (VR/AR): Foundations and Methods of Extended Realities (XR)* (pp. 1-37). Springer International Publishing. https://doi.org/10.1007/978-3-030-79062-2_1

Emerson, R. M., 1964. Power-Dependence Relations: Two Experiments. *Sociometry*, Vol. 27, No. (3), pp. 282–298. https://doi.org/10.2307/2785619

Felnhofer, A. et al., 2023. A Virtual Character's Agency Affects Social Responses in Immersive Virtual Reality: A Systematic Review and Meta-Analysis. *International Journal of Human–Computer Interaction*. https://doi.org/10.1080/10447318.2023.2209979

Fink, B. and Penton-Voak, I., 2002. Evolutionary Psychology of Facial Attractiveness. *Current Directions in Psychological Science*, Vol. 11, No. 5, pp. 154-158. https://doi.org/10.1111/1467-8721.00190

Freeman, G. and 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*, Vol. 4, Issue CSCW3, No. 239, pp. 1-27. https://doi.org/10.1145/3432938

Hatzakis, T., 2009. Towards a Framework of Trust Attribution Styles. *British Journal of Management*, Vol. 20, No. 4, pp. 448-460. https://doi.org/10.1111/j.1467-8551.2008.00596.x

Hewett, R. et al., 2018. Attribution theories in Human Resource Management research: A review and research agenda. *The International Journal of Human Resource Management*, Vol. 29, No. 1, pp. 87-126. https://doi.org/10.1080/09585192.2017.1380062

Homans, G. C., 1974. *Social behavior: Its elementary forms (Revised ed.)*. Harcourt Brace Jovanovich.

Jost, C. et al., (Eds.), 2020. *Human-Robot Interaction: Evaluation Methods and Their Standardization (Bd. 12)*. Springer International Publishing. https://doi.org/10.1007/978-3-030-42307-0

Kulms, P. and Kopp, S., 2019. More Human-Likeness, More Trust?: The Effect of Anthropomorphism on Self-Reported and Behavioral Trust in Continued and Interdependent Human-Agent Cooperation. *Proceedings of Mensch Und Computer 2019*, pp. 31-42. https://doi.org/10.1145/3340764.3340793

Kyrlitsias, C. and Michael-Grigoriou, D., 2022. Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey. *Frontiers in Virtual Reality*, Vol. 2. https://www.frontiersin.org/articles/10.3389/frvir.2021.786665

Montoya, R. M. and Horton, R. S., 2013. A meta-analytic investigation of the processes underlying the similarity-attraction effect. *Journal of Social and Personal Relationships*, Vol. 30, No. 1, pp. 64-94.

Natarajan, M. and Gombolay, M., 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 33-42. https://doi.org/10.1145/3319502.3374839

Raab, G., Unger, A. and Unger, F., 2010. Attributionstheorien. In G. Raab, A. Unger, & F. Unger (Eds.), *Marktpsychologie: Grundlagen und Anwendung* (pp. 77-96). Gabler. https://doi.org/10.1007/978-3-8349-6314-7_6

Rank, B., 1997. *Erwartungs-Wert-Theorien: Ein Theoriekonzept der Wirtschaftspsychologie und seine Anwendung auf eine berufsbiographische Entscheidung*. Rainer Hampp Verlag: München und Mering.

Rheu, M. (MJ), Jang, Y. and Peng, W., 2020. Enhancing Healthy Behaviors Through Virtual Self: A Systematic Review of Health Interventions Using Avatars. *Games for Health Journal*, Vol. 9, No. 2, pp. 85-94. https://doi.org/10.1089/g4h.2018.0134

Roebken, H., 2010. Similarity Attracts: An Analysis of Recruitment Decisions in Academia. *Educational Management Administration & Leadership*, Vol. 38, No. 4, pp. 472-486. https://doi.org/10.1177/1741143210368264

Sajjadi, P. et al., 2018. On the Effect of a Personality-Driven ECA on Perceived Social Presence and Game Experience in VR. *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1-8. https://doi.org/10.1109/VS-Games.2018.8493436

Santee, R. T., 1976. The Effect on Attraction of Attitude Similarity as Information about Interpersonal Reinforcement Contingencies. *Sociometry*, Vol. 39, No. 2, pp. 153-156. https://doi.org/10.2307/2786215

Seinfeld, S. et al., 2021. User Representations in Human-Computer Interaction. *Human–Computer Interaction*, Vol. 36, No. 5-6, pp. 400-438. https://doi.org/10.1080/07370024.2020.1724790

Seymour, M., Riemer, K. and Kay, J., 2018. Actors, Avatars and Agents: Potentials and Implications of Natural Face Technology for the Creation of Realistic Visual Presence. *Journal of the Association for Information Systems*, Vol. 19, No. 10, pp. 953-981. https://aisel.aisnet.org/jais/vol19/iss10/4

Seymour, M. et al., 2020. Facing the Artificial: Understanding Affinity, Trustworthiness, and Preference for More Realistic Digital Humans. *Proceedings of the 53rd Hawaii International Conference on System Sciences,* pp. 4673-4683. http://hdl.handle.net/10125/64316

Sherman, W. R. and Craig, A. B. (Eds.), 2020. Index of Media Experiences. In *Understanding Virtual Reality (Second Edition)*, pp. 903-908. Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-800965-9.18002-X

Simpson, J. A., 2007. Psychological Foundations of Trust. *Current Directions in Psychological Science*, Vol. 16, No. 5, pp. 264-268. https://doi.org/10.1111/j.1467-8721.2007.00517.x

Skalidis, I., Muller, O. and Fournier, S., 2022. CardioVerse: The cardiovascular medicine in the era of Metaverse. *Trends in Cardiovascular Medicine,* Vol. 33, No. 8, pp. 471-476. https://doi.org/10.1016/j.tcm.2022.05.004

Slater, M., Usoh, M. and Steed, A., 1994. Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, Vol. 3, No. 2, pp. 130-144. https://doi.org/10.1162/pres.1994.3.2.130

Sparwasser, P. et al., 2022. Virtual und Augmented Reality in der Urologie. *Der Urologe*, Vol. 61, No. 2, pp. 133-141. https://doi.org/10.1007/s00120-021-01734-y

Thomason, J., 2021. Metahealth-how will the metaverse change health care? *Journal of Metaverse*, Vol. 1, No. 1, pp. 13-16.

Vallverdú, J. and Trovato, G., 2016. Emotional affordances for human–robot interaction. *Adaptive Behavior*, Vol. 24, No. 5, pp. 320-334. https://doi.org/10.1177/1059712316668238

Weidner, F. et al., 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 29, No. 5, pp. 2596-2606. https://doi.org/10.1109/TVCG.2023.3247072

Witmer, B. G. and Singer, M. J., 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, Vol. 7, No. 3, pp. 225-240. https://doi.org/10.1162/105474698565686