PhD-FSTM-2023-119
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 14/11/2023 in Esch-sur-Alzette (Luxembourg)

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN INFORMATIQUE

by

Ningyuan SUN
Born on 16 April 1995 in Jining (China)

DELEGATION TO VIRTUAL AGENTS IN CRITICAL SCENARIOS: INFLUENCING FACTORS AND IMMERSIVE SETTINGS

Dissertation defence committee

Dr. Ulysses BERNARDET
Assistant Professor, Aston University

Dr. Jean BOTEV, Dissertation Supervisor
Research Scientist, University of Luxembourg

Dr. Pedro CARDOSO-LEITE, Vice Chairman
Associate Professor, University of Luxembourg

Dr. Steffen ROTHKUGEL, Chairman
Associate Professor, University of Luxembourg

Dr. Pieter SIMOENS
Associate Professor, Ghent University - imec

Abstract

Favored by the rapid advance of technologies such as artificial intelligence and computer graphics, *virtual agents* have been increasingly accessible, capable, and autonomous over the past decades. As a result of their growing technological prowess, interaction with virtual agents has been gradually evolving from a traditional user-tool relationship to one resembling interpersonal delegation, where users empower virtual agents to autonomously carry out specific tasks on their behalf. Forming a delegatory relationship with virtual agents can facilitate the user-agent interaction in numerous aspects, particularly regarding convenience and efficiency. Yet, it also comes with problems and challenges that may harm users drastically in critical scenarios and thus deserves extensive research. This thesis presents a thorough discussion of delegation to virtual agents based on a series of studies my colleagues and I conducted over the past four years. Several factors –including *agent representation*, *theory of mind*, *rapport*, and *technological immersion*– are examined individually via empirical approaches to reveal their impacts on delegation to virtual agents. A conceptual model featuring three interrelated dimensions is proposed, constituting a theoretical framework to integrate the empirical findings. An overall evaluation of these works indicates that users' decisions on delegating critical tasks to virtual agents are mainly based on rational thinking. Performance-related factors have a significant impact on delegation, whereas affective cues –such as rapport, agent representation, and theory of mind– are influential only to a limited extent. Furthermore, the usage of immersive media devices (e.g., head-mounted displays) has a marginal effect on users' delegatory decisions. Thus, it is advisable for developers to focus on performance-related aspects when designing virtual agents for critical tasks.

Acknowledgments

There are many people whom I would like to thank. Without their help and support, it would be terribly difficult –if not impossible– to finish my PhD.

I couldn't be more grateful for my advisor Jean Botev, who has always been a professional, open-minded, and patient mentor. An untold number of times, he gave me valuable feedback and advice, showing me the right way to go and saving me from consequential mistakes. Not just at work, my life in Luxembourg was also made easier and more colorful thanks to the many tips he offered.

I would like to thank my CET members, Pieter Simoens and Pedro Cardoso-Leite, for generously sharing their insights with me. I also want to express my appreciation to my coworkers, especially Steffen Rothkugel, Christian Grévisse, and Yara Khaluf.

Last but not least, many thanks to my wife Tianqing for her companionship and support throughout. Special thanks to our cats, Risotto and Poppi, who brought us so much joy.

My gratitude also goes to the Luxembourg National Research Fund (FNR) for funding my research in the context of the DELICIOS project (Delegation of Decision-Making to Autonomous Agents in Socio-Technical Systems, grant number 12635165).

Introduction

After decades of rapid development and proliferation, software programs have become integral to human societies. Their usage prevails a broad spectrum of activities, ranging from professional ones –such as 3D modeling, data analysis, and physics simulation– to daily utilities as in today's diverse applications for news, navigation, cooking, entertainment, and much more. The bulk of existing software programs are assistive, functioning mainly as a tool to enact users' strategies and initiatives. A typical example is word processors, through which users can edit and typeset the texts in a document through various commands and operations. However, without users' input, the program on its own can neither generate meaningful texts nor format them accordingly. Nevertheless, with the burgeoning development of artificial intelligence technologies, software programs are becoming more capable, autonomous, and proactive. As exemplified by ChatGPT, software programs can already automatically generate plausible texts in addition to their traditional role as a tool for editing and typesetting [OpenAI, 2023]. Researchers often refer to these intelligent software-based entities as *software agents* to differentiate them from tool-like software programs.

Interaction with software agents is similar to interpersonal delegation, where users empower software agents to carry out specific tasks on their behalf while users remain responsible for the task outcomes [Baird and Maruping, 2021; Milewski and Lewis, 1997]. Forming a delegatory relationship with software agents can benefit users in several aspects, particularly in terms of efficiency and convenience, as users can simply assign their tasks to software agents without the need to tender strategies and initiatives. On the other hand, this delegation-like interaction also entails challenges since users have to transfer their authority to software agents, making themselves vulnerable to the agents' actions. As more and more software agents are deployed into real-world scenarios, it is imperative to study this emerging delegatory relationship and identify its underlying factors – particularly when involving critical tasks whose outcomes may have far-reaching consequences– so as to facilitate the user-agent interaction and resolve potential issues therein.

Over the last four years, my colleagues and I have been studying users' delegation to *virtual agents*, a type of software agents that utilize and feature human communication –such as gesture, speech, and facial expression– to interact with users more naturally and personably than the traditional WIMP (windows, icons, menus, and pointer) interface allows [Lugrin, 2021]. While the term "virtual agent" often refers to software agents embodied as virtual characters (e.g., virtual humans), it also includes those represented by textual interfaces (e.g., ChatGPT) to emulate interpersonal conversations. Either way, their advanced communication capabilities were frequently demonstrated to impact user behaviors [Lugrin et al., 2021, 2022] and, therefore, may also influence user decisions on delegation. Our research focuses on two relevant but understudied aspects:

1. Influencing factors in delegation decisions to virtual agents.
2. Delegation to virtual agents in immersive settings.

In light of our initial findings detailed in Chapter 5, we concentrated on embodied virtual agents, especially those with humanlike representation such as virtual humans or robots. Through nine chapters, this thesis presents and integrates the theoretical and empirical studies we conducted, striving to provide new insights into the two topics. The first two chapters give a bird's eye view of the literature on delegation (cf. Chapter 1) and virtual agents (cf. Chapter 2), while the remaining chapters are arranged into the following thematic blocks.

Theoretical Foundation

Numerous factors may influence delegation to virtual agents, making up a huge parameter space that remains largely unknown due to scant research on this topic. Chapter 3 presents an overview of potentially relevant factors and introduces a model conceptualizing how these factors collectively shape users' trust in and delegation to virtual agents. Taking a macroscopic perspective, the model differentiates three major dimensions covering the impact of rationality, affection, and technology on delegation. The model serves as the theoretical foundation to connect and integrate the empirical studies detailed in the subsequent chapters.

Empirical Research

A large part of the thesis (Chapters 4 to 8) discusses our empirical studies investigating different factors' impact on delegation to virtual agents.

Informativeness (i.e., the extent to which an agent provides useful and sufficient information) may constitute a relevant factor in delegation to virtual agents, according to the exploratory study detailed in Chapter 4. Through an online survey, participants ($N = 80$) were asked to rank the importance of 20 factors in their decisions on delegation to virtual agents. Informativeness was rated the most salient factor early during the interaction. Its initial influence on delegatory decisions is potentially stronger than those of other acknowledged factors such as capability or trustworthiness.

The *visual representation* of virtual agents was also found to be a relevant factor in another survey-based study presented in Chapter 5. Participants ($N = 100$) were tasked to rank five categories of agent representation with varied levels of visual human-likeness. The results delineated a dividing line between the preference for humanlike (humans and humanoids) and non-human representations (organic non-human, inanimate objects, and symbols), where humanlike representations were preferred over non-human ones.

Theory of mind (here defined as an agent's ability to infer others' beliefs) showed a limited impact on delegatory decisions in the between-subjects experiment discussed in Chapter 6. Participants ($N = 75$) were divided into three groups, each interacting with a virtual agent that has a unique level of theory of mind. The results indicated a non-monotonic relationship between the agents' theory of mind level and participants' delegatory decisions or intentions. This contradicts the hypothesis that users are more likely to delegate to virtual agents with higher-level theory of mind.

Performance, particularly users' performance or the difference between users' and virtual agents' performance, is an essential factor in delegating to virtual agents. The prominence of performance was demonstrated in the experiment on theory of mind (cf. Chapter 6) –where performance and delegatory intention were negatively correlated– and reflected in the other studies throughout the thesis (e.g., Chapters 4 and 8).

Technological immersion (i.e., a media device's capacity to deliver immersive experiences) may only be an insignificant factor in users' decisions on delegation, as discussed in Chapter 7. Two experiments (both $N = 30$) employing different designs were conducted to observe and compare users' delegatory behavior toward virtual agents in varied conditions and levels of technological immersion. Contrary to the hypothesis that technological immersion is a relevant factor, the results showed that participants' decisions on delegation were similar across the different conditions, indicating its marginal influence.

Rapport, albeit being important for interpersonal communication, may only play a lesser role in user-agent interaction involving decisions on delegation to virtual agents. Chapter 8 elaborates on a between-subjects experiment ($N = 15$) that contrasts delegation to a rapport-building and rapport-avoiding agent in virtual reality. The results gave an initial indication that perceived rapport may influence users' delegatory decisions only to a limited extent.

Integration, Reflection, and Outlook

The presented work is recapped and integrated in Chapter 9, where the influence of each dimension in the conceptual model on delegation is evaluated based on the results of the empirical studies.

The *analytical dimension* is the most relevant due to the considerable impact of performance and informativeness on users' decisions on delegation to virtual agents. The *affective dimension* is influential to some extent given the mixed evidence: visual representation is relevant, yet theory of mind and rapport may not be significant factors. The *technological dimension* may only have a limited effect since none of the evidence collected so far can validate the influence of technological immersion on delegation.

Several guidelines for virtual agent design can be derived from the findings. For example, since the analytical dimension is the most relevant, developers should focus on performance-related aspects when designing virtual agents to carry out critical tasks autonomously. A limited number of affective cues (e.g., facial expression, gesture) can make the interaction more natural and personable. However, adding more affective cues is unlikely to yield better outcomes since the affective dimension is influential only to some extent. Excessive affective cues may even backfire if poorly designed, as exemplified by the notorious Uncanny Valley effect. As for the choice of media device for virtual agents, developers should mainly consider the nature of the task rather than how media devices influence virtual agent perception.

Finally, the limitations of our methodologies are also discussed from a macroscopic perspective, and future research avenues are outlined at the end to conclude this thesis.

Table of Contents

List of Abbreviations

AR	Augmented Reality
IS	Information System
SD	Standard Deviation
ToM	Theory of Mind
VR	Virtual Reality
WIMP	Windows, Icons, Menus, Pointer

Chapter 1

An Overview of Research on Delegation

Delegation is one of the most important and common interpersonal relationships in human societies. The operation of most organizations requires effective management, i.e., delegation from managers to subordinates, and successful collaboration, i.e., delegation among specialized members or departments within an organization. Even God delegates, as noted in the *Book of Exodus*: “Behold, I send an Angel before thee, to keep thee in the way, and bring thee into the place which I have prepared.” Due to its significance and prevalence, delegation has been the subject of much debate across several disciplines as diverse as economics, management, and politics [Lupia, 2001]. Recent years have also seen the emergence of a new research line on delegation to intelligent artifacts, such as robots and software agents [Baird and Maruping, 2021]. This chapter gives an overview of these studies in two folds. Section 1.1 introduces a widely accepted theory and different factors of interpersonal delegation, followed by Section 1.2 overlooking the development of and recent work on delegation to software agents.

1.1 Interpersonal Delegation

Delegation is often regarded as a dyadic relationship between a *principal* and an *agent*, where the principal lets the agent carry out specific tasks on the principal's behalf while the principal remains responsible or accountable for the task outcomes [Leana, 1987; Milewski and Lewis, 1997]. Delegation can provide principals with several benefits, a common one of which is allowing principals to access and utilize agents' expertise and knowledge [Jensen, 1986]. Through delegation, principals can also increase their working efficiency by shifting some of their workloads to agents [Yukl and Fu, 1999]. Principals may even use delegation strategically as a commitment device to express their determination in negotiation [Sengul et al., 2012]. On the other hand, principals must bear the risk of transferring part or all of their authority to agents during delegation. Such risk can be even higher when principals lack effective measures to know and regulate their agents' actions [Eisenhardt, 1989].

One of the most influential works on delegation is *agency theory* [Eisenhardt, 1989; Jensen and Meckling, 1976; Shapiro, 2005]. It originates from early studies on risk-sharing problems within an organization whose members hold different attitudes toward risk [Wilson, 1968]. Later, these studies evolved into a more inclusive theory, i.e., agency theory, mainly addressing the delegatory relationship between two rational entities. The theory posits that delegation is susceptible to the so-called *agency problems*, which often arise in two situations: (a) when it is difficult for principals to verify what agents are actually doing, and (b) when principals and agents have conflicting goals or different attitudes toward risks [Eisenhardt, 1989]. These problems are detrimental to delegation, as exemplified by the so-called *moral hazard*, a situation where agents do not act as agreed or even furtively impair principals' interests [Shapiro, 2005]. To resolve these problems, agency theory advocates to use a set of rules –i.e., a *contract*– to regulate the principal-agent dyad, so as to minimize the negative influence emanating from the agency problems and maximize the dyad's efficiency [Eisenhardt, 1989; Jensen and Meckling, 1976]. There are mainly two types of contracts. With *behavior-oriented* contracts, agents receive a fixed amount of money or something equivalent upon finishing the tasks delegated to them, regardless of the task outcomes. *Outcome-oriented* contracts give agents minimal remuneration for finishing the tasks, but they can take a share of the task outcomes (e.g., bonus, commission, stocks). The two types of contracts fit different situations depending on several factors, including agent observability [Fama, 1980; Fama and Jensen, 1983], task programmability [Eisenhardt, 1985, 1988], the measurability of task outcomes [Anderson, 1985], and more. For example, in programmable tasks, principals are more likely to employ behavior-oriented contracts because agents' performance in such tasks is relatively easy to observe and assess [Eisenhardt, 1989].

A large part of the literature on delegation is related to agency theory, viewing the principal-agent relationship mainly as an economic issue [Alonso and Matouschek, 2007; Grossman and Hart, 1992; Ross, 1973]. Nevertheless, there are also many researchers who assume a psychological stance and focus on principals' decision-making process of delegation. For example, Leana [1987] identified eight factors underlying a manager's decisions on delegating a task to subordinates, including the manager's workload, the task importance, and subordinates' age, gender, trustworthiness, performance, job capability, and job tenure. This set of factors was later validated in broader contexts and expanded also to include subordinates' experience in managing, the goal congruence between managers and subordinates, and whether there is a favorable exchange relationship [Aggarwal and Mazumdar, 2008; Yukl and Fu, 1999]. In some cases, decisions on delegation are dominated by subjective causes. For example, people may delegate a task to others to avoid the feeling of being responsible or blamed for negative task outcomes [Steffel and Williams, 2018; Steffel et al., 2016]. A manager may refuse to delegate to appear "busy" [Jenks and Kelly, 1985]. Overall, these psychology-oriented studies suggest that delegation is a challenging and prone-to-failure task for principals. Managers of all levels, from leaders of small groups to CEOs of successful corporations, are not free from making sub-optimal or wrong decisions on delegation [Jenks and Kelly, 1985]. People generally prefer making decisions themselves over delegating them to others, even if retaining control would lose potential benefits and incur additional costs [Bobadilla-Suarez et al., 2017].

1.2 Delegation to Software Agents

While the term "delegation" has been predominantly used and studied for interpersonal contexts, recent years have seen the term gradually adopted in computer science research as interaction with software agents becomes increasingly similar to delegation. Unlike traditional software programs that rely heavily on users' strategies and initiatives, software agents are more autonomous and capable of performing complicated tasks where the goal is ambiguous (e.g., providing psychotherapy to users) or where the external environment is highly dynamic and imbued with uncertainties (e.g., stock exchange). In the literature, interaction with software agents has been primarily studied from the perspectives of use [Parasuraman and Riley, 1997; Venkatesh et al., 2003] and reliance [Dixon and Wickens, 2006; Riley, 1996]. As software agents become increasingly autonomous and capable, a different perspective based on the notion of delegation recently received more attention, as reflected in the quote below.

> [...] the delegation lens will yield more relevant and nuanced insights regarding human agent and agentic IS[1] artifact relationships, and this lens will be increasingly needed as the agentic capabilities of IS artifacts increase. [Baird and Maruping, 2021]

One of the pioneering works on delegation to software agents can be traced back to 1997 when Milewski and Lewis [1997] raised their concerns about the "dramatic change" of the software user interface from tool-oriented to delegation-oriented ones. They argued that, as a result of the rapid change, problems common to interpersonal delegation might similarly occur during interaction with software agents, which imposed new challenges on user interface designs. "To overcome well-established drawbacks in delegation", they listed five major dimensions to consider when designing delegation-oriented user interfaces, including trust, communication, performance control, users' demographics, and cost-benefit analysis.

Another early study formalized delegatory relationships within a multi-agent system [Castelfranchi and Falcone, 1998]. Although targeting inter-agent instead of user-agent delegation, their theory is still interesting in a broader human-computer interaction context since it offers a unique perspective on delegation to software agents. Different from other approaches, Castelfranchi and Fal-

---

[1] IS = information systems

cone [1998] regarded delegation as a state of principals, where "an agent A needs or likes an action of another agent B and includes it in its own plan". In this definition, they removed the element of responsibility and accentuated principals' demand or preference for specific actions from agents. Following the definition, they classified delegation into three categories (cf. Table 1.1) based on two criteria: whether there is an agreement between a principal and an agent and whether the principal actively induces certain behaviors in the agent.

Table 1.1: The delegation classification in [Castelfranchi and Falcone, 1998].

Delegation type	**Agreement exists?**	**Behavior inducing?**
Weak delegation	No	No
Mild delegation	No	Yes
Strict delegation	Yes	Yes or No

After these pioneering works, delegation to software agents as a concept has been sporadically explored, with some relevant studies emerging only in recent years. A frequently discussed topic is whether people prefer human agents or software agents. The evidence is mixed; some studies found that software agents are preferred [Candrian and Scherer, 2022; Fügener et al., 2022; Logg et al., 2019], whereas others showed the opposite [Dietvorst et al., 2015]. Many factors can influence this preference. For example, people may prefer letting software agents carry out tasks involving sensitive data (e.g., credit card information) due to their user-centered design [Fogg, 2009; Sundar and Kim, 2019] and limited intentional capacity [Harbers et al., 2017; Sundar and Kim, 2019], whereas human agents may exploit the sensitive data for their own interests. On the other hand, tasks involving moral decisions (e.g., life-and-death decisions in law, military, or medicine) are less likely delegated to software agents given their lack of empathy [Bigman and Gray, 2018; Liehner et al., 2021]. Lubars and Tan [2019] conceptualized the term *delegability* referring to people's general preference of delegating a task to artificial intelligence. They conducted a survey on the delegability of 100 different tasks, of which the ones with the highest and lowest level of delegability were "moving & packing merchandise in a warehouse for shipping to customers” and “picking out and buying a birthday present for an acquaintance”, respectively.

Another often discussed topic is the factors governing users’ decisions on delegation to software agents. Research shows that many factors have similar impacts on delegation to human and software agents. To give a few examples, interpersonal and human-software delegation were both found to be positively correlated with perceived controllability [Stout et al., 2014], perceived attachment [Leyer et al., 2021], and agents’ trustworthiness [Lubars and Tan, 2019; Stout et al., 2014]. Notably, perceived controllability plays an essential role in delegation to software agents; people are more likely to delegate a decision to an algorithm when they are allowed to modify the decision made by the algorithm, even if the modification is severely restricted [Dietvorst et al., 2018]. There are also factors whose effects on interpersonal and human-software delegation are opposite. High-level task accountability, for instance, can encourage delegation to software agents but inhibit delegation to human agents [Stout et al., 2014].

Several studies investigated delegation to software agents from an economic perspective and focused on the user-agent dyad [Fernández Domingos et al., 2022; Fügener et al., 2022, 2021; Hukal et al., 2019]. An interesting implication from these studies is that software agents may be more proficient at delegation than humans. In certain tasks, a hybrid team of a human user and a software agent can reach peak performance when the software agent assumes the leading role and delegates tasks to the human user [Fügener et al., 2022]. In contrast, the team’s performance is relatively low when the human user is the leader or when any party seizes full control [ibid.].

1.3 Chapter Summary

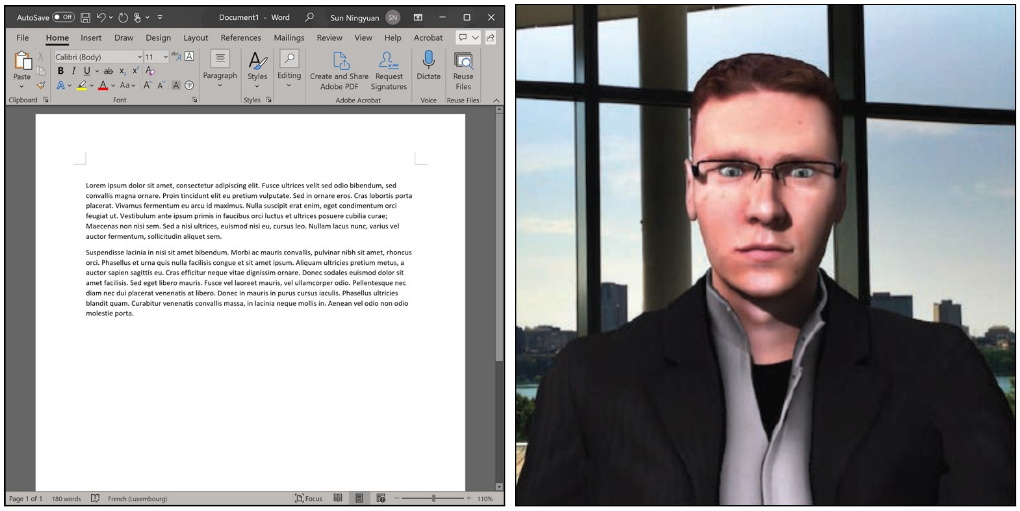
This chapter provided an overview of research on delegation, starting with the extensively studied topic of interpersonal delegation, then moving on to the recently emerged line of research on delegation to software agents. These studies yielded many insights regarding delegation, among which three points deserve the attention of software agent developers: (1) in delegation, people are vulnerable to agents' actions due to the transfer of authority; (2) people tend to retain control of a task rather than delegate the task to others; (3) people are not innately good at making the right decision on delegation. These issues may similarly occur during the increasingly delegation-like interaction with software agents, impairing the user experience and lowering the overall efficiency. To address these issues, developers may consider using virtual agents, a type of software agent that utilizes human communication –such as facial expression and gesture– to interact with users in a natural and personable fashion. With their ability to leverage various verbal and non-verbal behaviors, virtual agents have the potential to establish an efficient and satisfying delegatory relationship with users. The next chapter briefly introduces virtual agents in several aspects, covering their design, applications, and limitations.

Chapter 2

A Brief Introduction to Virtual Agents

Since the advent of the graphical user interface, interaction with software programs has been predominantly mediated by graphical widgets such as windows, menus, icons, and pointer, i.e., the so-called WIMP interface (cf. Figure 2.1a). Not only in daily applications, its pervasive use can also be seen in the literature on delegation to software agents, where most empirical studies (cf. Section 1.2) employed the WIMP interface to mediate the interaction between human participants and software agents. While the WIMP interface remains the common practice today, there has been lively discussion and research looking for more natural and personable ways to interact with users than the WIMP interface allows. The prevailing approach is to enable software agents to use human communication, such as speech or facial expression, to complement or even replace the WIMP interface. These agents are commonly called *virtual agents*, as exemplified by digital humans (cf. Figure 2.1b) and the rapidly proliferating generative artificial intelligence (e.g., ChatGPT).

This chapter gives a brief introduction to virtual agents, mainly covering embodied virtual agents that this thesis concentrates on. Section 2.1 flies over several design dimensions of virtual agents, followed by Section 2.2 and Section 2.3 listing their applications and limitations, respectively. Section 2.4 specifically discusses virtual agents experienced in immersive settings.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce ultrices velit sed odio bibendum, sed convallis magna ornare. Proin tincidunt elit eu pretium vulputate. Sed in ornare eros. Cras lobortis porta placerat. Vivamus fermentum eu arcu id maximus. Nulla suscipit erat enim, eget condimentum orci feugiat ut. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Maecenas non nisi sem. Sed a nisi ultrices, euismod nisi eu, cursus leo. Nullam lacus nunc, varius vel auctor fermentum, sollicitudin aliquet sem.

Suspendisse lacinia in nisi sit amet bibendum. Morbi ac mauris convallis, pulvinar nibh sit amet, rhoncus orci. Phasellus et urna quis nulla facilisis congue et sit amet ipsum. Aliquam ultricies pretium metus, a auctor sapien sagittis eu. Cras efficitur neque vitae dignissim ornare. Donec sodales euismod dolor sit amet facilisis. Sed eget libero mauris. Fusce vel laoreet mauris, vel ullamcorper odio. Pellentesque nec diam nec dui placerat venenatis at libero. Donec in mauris in purus cursus iaculis. Phasellus ultricies blandit quam. Curabitur venenatis convallis massa, in lacinia neque mollis in. Aenean vel odio non odio molestie porta.

(a) A WIMP interface. **(b)** A virtual agent.

Figure 2.1: Examples of a WIMP interface and a virtual agent. Figure 2.1a is a screenshot of Microsoft Word whose user interface is largely WIMP-based. Figure 2.1b (from [Hoque et al., 2013]) depicts a virtual conversational coach that users can converse with for social skill training.

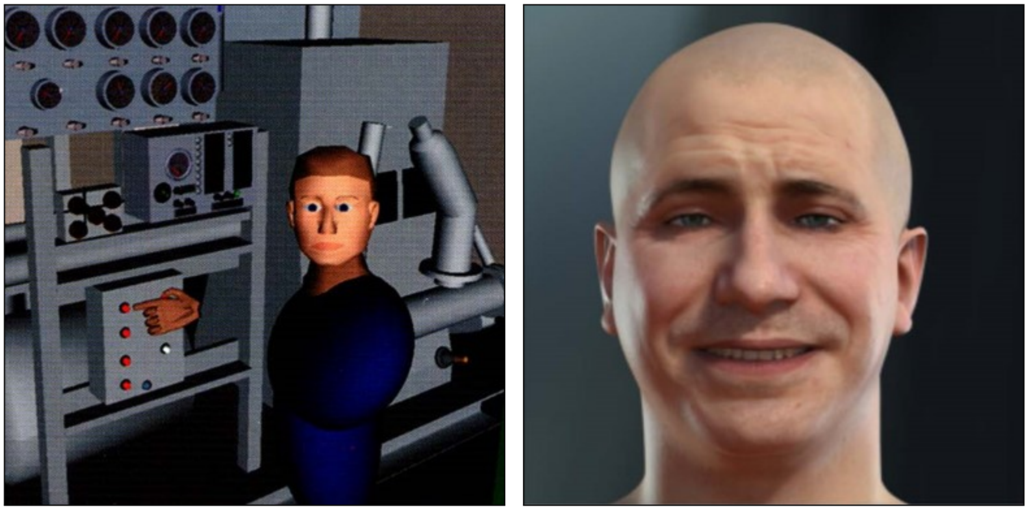
2.1 Design Dimensions

Virtual agents often have a virtual body that allows them to communicate with users using non-verbal behaviors. The presence of a virtual body can yield various positive impacts on user-agent interaction, such as improving the perception of virtual agents [Koda and Maes, 1996; Rosenberg-Kima et al., 2008], making the interaction more engaging [Koda and Maes, 1996], and increasing the naturalness of communication [Sproull et al., 1996], the sense of social presence [Kim et al., 2018], and agents' trustworthiness [Weitz et al., 2019]. However, the virtual body may also produce adverse effects, for instance, when it triggers the notorious Uncanny Valley phenomenon [Mori, 1970] or contains culturally offensive elements.

The virtual body can be visually represented in a plethora of forms. Merely human representation alone already varies in a multitude of details, such as age, gender, ethnicity, body figure, clothes, and facial attractiveness, not to mention other categories of representation like animals, humanoids, or inanimate objects. Each detail potentially has its unique impact on user-agent interaction; for example, a virtual agent with human representation may appear more trustworthy when it is embodied in a female character than a male one [Surprenant, 2012] or when it wears light-color clothes than dark-color ones [Peña and Yoo, 2014]. However, compared to the vast possibilities in agent representation, little is known regarding their implications, and the literature today can only provide limited insights about using certain general categories of representation.

As computer graphics technologies rapidly advance, the visuals of virtual agents have been constantly evolving in terms of qualities, art styles, production pipelines, and other aspects (cf. Figure 2.2). The significant progress offers virtual agent developers immense possibilities and latitudes when designing the graphics of agent representation. Evidence shows that the perception of a virtual agent is contingent on the graphical style of its visuals [McDonnell et al., 2012]. High-level visual realism (e.g., photo- or hyper-realism) can increase the acceptance of a virtual agent [Ring et al., 2014; van Wissen et al., 2016] or the sense of presence [Hvass et al., 2017] but may risk falling into the Uncanny Valley. Virtual agents with a less realistic visual representation (e.g., cartoon or stylized ones) are perceived as more friendly [Robertson et al., 2015] and appealing [McDonnell et al., 2012] yet may appear inappropriate for critical scenarios such as medical information communication [Ring et al., 2014; Robertson et al., 2015].

Apart from its static appearance, the virtual body can also be animated to emulate non-verbal behaviors, which constitute a staple part of interpersonal face-to-face communication. They can express a rich repertoire of meanings [Berry et al., 1997; Mehrabian and Ferris, 1967; Mehrabian and Wiener, 1967]. For example, a smile can indicate radically different messages depending on who is smiling (e.g., a mother vs. politician) or where the smile occurs (e.g., at a birthday party vs. a funeral). Smiles of nuanced differences can be interpreted as diverse as signals of amusement, politeness, or embarrassment [Ochs et al., 2010]. In addition to the rich meanings they convey, non-verbal behaviors can also be more efficient than verbal or textual communication in certain cases, such as those involving spatial tasks (e.g., pointing to an object using a finger than describing its position using words) [Harper et al., 1978]. Moreover, unlike



(a) A virtual agent ca. 1997.

(b) A virtual agent ca. 2019.

Figure 2.2: Evolution of computer graphics in research on virtual agents. Figure 2.2a (from [Johnson and Rickel, 1997]) depicts Steve, a pedagogical agent developed before the millennium. Its crude graphics were typical in studies on virtual agents during that time. Figure 2.2b (from [Torre et al., 2019]) shows an agent developed in recent years. Its graphics surpass Steve's in order of magnitudes regarding texture resolution, light tracing, and other aspects.

natural languages which require minimal levels of learning before being used, non-verbal communication is readily accessible to most people, regardless of their demographics or cultural backgrounds. Certain non-verbal behaviors are universal and can serve as an alternative or complement when verbal communication is difficult. When integrated into virtual agents, non-verbal behaviors can exert various influences on human-agent interaction. To give some examples, the list below summarizes the impacts of several commonly studied non-verbal behaviors in the literature.

Facial Expression

- Students can achieve better learning outcomes after hearing a lecture given by a virtual agent with relatively rich facial expressions than a virtual agent whose facial expression stays neutral [Baylor and Kim, 2009; Theonas et al., 2008].

- Children are more likely engaged in stories narrated by virtual agents with expressive facial expressions than with neutral ones [Şerban et al., 2017].
- When a human player plays the *dictator game* against a virtual agent, the player will donate more money if the agent exhibits more dynamic facial expressions [Takagi and Terada, 2021].
- Users are more likely to feel a rapport with virtual agents showing positive facial expressions than those with negative ones [Wong and McGee, 2012].
- Trainees perform better in the programs hosted by a virtual agent with a happy face than an agent that looks angry [Garcia et al., 2016].

Gesture

- A virtual agent's gestures are relevant to users' perception of its extroversion [Neff et al., 2010], likability [Bergmann et al., 2010], human-likeness [ibid.], and competence [Bergmann et al., 2012, 2010].
- For an embodied conversational agent, making a self-touching gesture (i.e., touching itself with its hands) renders the agent more warmhearted, agile, committed, and natural, but more aggressive and strained [Krämer et al., 2007].
- In lectures given by virtual agents, students are more likely to be engaged when the agents utilize gestures in their teaching, but their gestures may also be distracting and decrease teaching outcomes [Davis and Antonenko, 2017].
- When a virtual agent tries to explain a software program to users, its gestures can modulate the perceived quality of its explanation [Buisine et al., 2004].

Gaze

- Virtual agents mimicking human gazing patterns are perceived more positively regarding ease of use [Heylen et al., 2002] and naturalness [ibid.].
- Random gazing behaviors are generally detrimental to users' perception of virtual agents [Garau et al., 2001].
- The frequency of an agent's direct gaze (i.e., looking at users) and averted gaze (i.e., not looking at users) can modulate the perception of the agent's social

dominance and users' stress [Obaid et al., 2012], engagement [Andrist et al., 2012; Bee et al., 2009], and learning outcomes [Andrist et al., 2012].

Body Movement

- Virtual agents showing relatively open postures are perceived as having a higher level of willingness to interact with others [Li et al., 2018].
- Virtual agents mimicking users' body movements are rated more positively in terms of rapport [Raffard et al., 2018], trustworthiness [Aburumman et al., 2022], likability [ibid.], and social closeness [Tarr et al., 2018].
- Gait features, such as the degree of arm swings, can be used in virtual agents to modulate their emotional expressivity [Randhavane et al., 2019].

Virtual agents' non-verbal behaviors are mainly emulated by replaying pre-recorded animation clips; for example, an embodied conversational agent can trigger a head-nodding animation clip as a signal of backchanneling when it detects a pause in users' speech [Gratch et al., 2006]. This approach is suitable for non-verbal behaviors made of simple motions such as eye blinks. When it comes to more complex ones like facial expressions, the pre-recording approach is plausible but falls short since the dynamics and varieties of these behaviors are often too complicated to be fully covered by pre-recordings. A promising solution is to use *autonomous animation*, either through procedural or data-driven techniques, so that virtual agents can partially or fully manage the animation with autonomy [Klein et al., 2019; Perlin and Seidman, 2008]. Autonomous animation has been successfully applied to non-human characters (e.g., animating a spider's legs) but remains challenging for human characters.

2.2 Applications

The application of virtual agents can be seen in many fields, among which education is a frequently discussed scenario. Early *pedagogical agents* (i.e., virtual agents for pedagogical purposes [Lane and Schroeder, 2022; Schroeder et al., 2013]) can be dated back to the last century when Johnson and Rickel [1997] created the virtual agent Steve to teach students procedural tasks like operating a complex machine. Like Steve, many pedagogical agents are designated the role of a teacher to provide users with guidance and feedback. Learning through or with pedagogical agents can increase students' self-efficacy, motivation, and engagement [Lane and Schroeder, 2022]. Moreover, lessons with pedagogical agents can be arranged flexibly at students' convenience or repeated multiple times without incurring additional costs.

Apart from acting as a teacher, virtual agents can also assume the role of classmates to facilitate education. Peer-based learning, such as group work, is a widely acknowledged and frequently used teaching method today [Blum-Kulka and Dvir-Gvirsman, 2010; Porter et al., 2013; Sin et al., 2019]. Evidence shows that learning with virtual peers is similarly effective as with fellow students in terms of learning outcomes [Cassell, 2004] and motivation [Lane et al., 2013]. The value of virtual peers can be more prominent in special periods like the COVID-19 pandemic, during which meeting with fellow students was difficult due to sanitary measures, whereas virtual peers are always accessible. Not only in education, peer-like virtual agents are also used in other contexts where social connection is desirable. For instance,they may offer company for older adults who live alone [Cho, 2021] or people with cognitive impairment [Demiris et al., 2016] as an alternative to human companions. Many video games include virtual agents as players' teammates (e.g., computer-controlled allies in strategic games) to improve engagement and social presence [Prada and Rato, 2022].

In healthcare domains, virtual agents may persuade individuals to change their unhealthy lifestyles (e.g., substance abuse [Olafsson et al., 2020b; Yasavur et al., 2014]) and adopt better ones (e.g., exercising regularly and eating nutritiously [Olafsson et al., 2020a]). In communication with patients, virtual agents can be equally effective as doctors in terms of warmth and competence [Dai and MacDorman, 2018]. Patients may even prefer virtual agents over hospital personnel for communication due to their low levels of health literacy (i.e., the ability to read and understand healthcare information) [Zhou et al., 2014] or because they do not feel rushed when interacting with virtual agents, whereas doctors may appear in a hurry to finish the conversation [ibid.].

Researchers often use virtual agents to simulate real-world scenarios (e.g., public speech) to study human behaviors [Bombari et al., 2015; Pan and Hamilton, 2018]. In experiments that adopt this methodology, participants are usually tasked to interact with one or multiple virtual agents whose behaviors follow a pre-defined script to create a social context [Latu et al., 2013]. Evidence shows that people react to computer-simulated social signals in similar ways as they would in real interpersonal interactions [Nass et al., 1994]. There are several advantages to using virtual agents than employing human actors. Virtual agents are computer-generated and thus are fully controllable to researchers in terms of their appearance and behaviors. Virtual agents can also perfectly replicate their behaviors when interacting with different participants. In contrast, human actors are hardly able to sustain their behavioral consistency to the same level, which consequently may decrease the validity of experiment results.

2.3 Limitations

Developing a sufficiently robust and interactive virtual agent is a complex task requiring considerable time and effort, as well as multidisciplinary expertise in art, animation, programming, psychology, and more [Lane and Schroeder, 2022, p. 308]. Although increasingly powerful tools like the Nvidia Omniverse have streamlined the creation pipeline, individual developers may still find themselves devoid of required skills or resources and, consequently, can only produce virtual agents of limited capabilities. Moreover, several key technologies for virtual agents (e.g., autonomous animation, non-verbal communication) are arguably still immature to unlock their full potential.

In many circumstances, virtual agents are not the ideal interface to mediate the interaction with users. Virtual agents are characterized by their capabilities of human communication and thus are suitable for tasks that require maintaining a social bond with users. However, when the bond is marginally relevant, their capabilities of human communication may bear extra cognitive load on users [Clark and Choi, 2005; Lugrin, 2021]. Furthermore, the virtual-agent interface is not the panacea for all social scenarios, either. In quite some cases, people prefer interacting with *social robots*, which share the same physical space with users and thus afford a stronger sense of physical and social presence [Li, 2015; Thellman et al., 2016]. Although devices like VR headsets allow users and virtual agents to share the same virtual space, these devices are still developing to mediate a virtual environment as immersive as reality.

Virtual agents are inherently simulated, so there is immense latitude in their design. This great liberty offers many advantages, but it can also be problematic if handled carelessly or exploited. Racial and gender prejudice may occur when the agent representation resembles individuals who are more frequently subjected to discrimination than others [Cassell, 2022, p. 353]. Certain forms of representation are considered disrespectful of one's religious belief. Virtual agents in video games are often imbued with violent or sexual elements, the exposure of which was demonstrated to have negative impacts on players [Dowsett and Jackson, 2019; Guggisberg, 2020]. In extreme cases, people may become obsessed with interacting with virtual agents and detached from real-world social lives. Such obsession can be manifested as video game addiction or extreme virtual character fandom [Cassell, 2022, p. 354].

2.4 Virtual Agents in Immersive Settings

Interacting with virtual agents in immersive settings has been increasingly common due to the proliferation of *immersive media devices*, such as consumer-grade VR and AR headsets. These devices are usually designed to provide realistic perceptual cues –e.g., stereoscopic images, panoramic field of regard, high-resolution tracking, and more– to envelop users in a mediated environment as if users were being there in the environment or as if virtual objects were being there in reality. Research shows that immersive media devices afford a higher level of spatial and social presence than commonly used ones like desktops, laptops, smartphones, and tablets [Buttussi and Chittaro, 2018; Cummings and Bailenson, 2016; Kim et al., 2014; Makransky et al., 2019; Moreno and Mayer, 2004; Ochs et al., 2019; Reinhard et al., 2022]. The heightened sense of presence can further impact user-agent interaction by, for example, increasing agents' believability [Hepperle et al., 2022; Ochs et al., 2019] or making agents' emotions appear more expressive [Faita et al., 2016].

Deploying virtual agents in immersive settings allows developers to utilize the potential of haptics-based interaction and facilitate its implementation. Gloves or sleeves with vibrotactile feedback are not uncommon in VR settings for simulating the sense of touch [Boucaud et al., 2019; Gallace and Girondini, 2022]. Certain wearable devices can provide finer haptic cues related to temperature [Cai et al., 2020] or textures [Cai et al., 2020; Keef et al., 2020]. Simulated *social touch*[1] with virtual agents in immersive settings can improve the perception of virtual agents, including their warmth [Huisman et al., 2014], empathy [Boucaud et al., 2019], and closeness [Hoppe et al., 2020], as well as increasing perceived agency [ibid.] and co-presence [ibid.]. A social touch from virtual agents may even change users' behaviors directly, as exemplified by the so-called *Midas touch effect* [Zhao et al., 2018].

Another notable topic here pertains to the *Proteus effect*, which states that people's behavior in a digital world (video games, online chat rooms, etc.) can be impacted by the appearance of their avatars or self-representation [Yee and Bailenson, 2007]. For example, when negotiating with another individual in a shared virtual environment, people will behave more confidently if their avatars are taller than the other's avatar, regardless of their actual height [ibid.]. Many VR applications by default represent users in avatars, often as a full-body human, a half-body human with the upper part only, or simply a pair of floating hands. The frequent use of avatars makes the Proteus effect a non-negligible factor for user-agent interaction in immersive settings, particularly given the evidence that the Proteus effect tends to be stronger in immersive settings [Sakuma et al., 2023].

Despite their many potentials for user-agent interaction, immersive media devices are still fettered by some problems, of which a prominent one lies in their ergonomics. Compared to laptops and smartphones, immersive media devices often involve more exhausting body movements and thus can easily inflict cybersickness [Caserman et al., 2021] or fatigue [LaViola Jr. et al., 2017, p. 73] on users. Also, sensors embedded in these devices are usually more invasive to

---

[1] Social touch –i.e., touching or being touched by others, such as hugging, pushing, or kissing– can exert various influences on an individual, ranging from pleasant ones like a tender hug to negative ones as in the case of being jostled by others on a crowded bus.

achieve high-resolution tracking, which, on the other hand, may impose a higher risk on users' personal privacy and data security [Adams et al., 2018].

2.5 Chapter Summary

This chapter presented a brief introduction to virtual agents, touching upon their design (e.g., visual representation, graphical styles, non-verbal behaviors, and animation), applications (e.g., education, healthcare, and academia), limitations (e.g., resource-intensive development, high cognitive load, ethical concerns), and combination with immersive media technologies (e.g., altered perception, social touch, ergonomic issues, and cybersickness). The literature on virtual agents is vast. For a more comprehensive overview of related research, readers can refer to [Lugrin et al., 2021] and [Lugrin et al., 2022].

The next chapter introduces a conceptual model for trust in and delegation to virtual agents. The model considers numerous relevant factors and provides an explanation of how these factors collectively shape users' trust in and delegation to virtual agents.

Chapter 3

A Conceptual Model for Trust in and Delegation to Virtual Agents

Portions of the content presented in this chapter were published in [Sun and Botev, 2021a].

Trust is an essential commodity of almost any individual, institutional, or organizational relationship. Due to its importance and omnipresence, trust has been extensively studied in many disciplines, particularly economics, psychology, and sociology. Although the majority of these studies focus on interpersonal trust, recently there emerged a large body of research on the trusting relationship between users and intelligent artifacts such as robots, automation, and software agents. Numerous factors were identified and demonstrated to impact users' trust in these artifacts, constituting a vast and complicated parameter space.

Delegation to virtual agents is potentially governed by a similar parameter space, which, unlike that of trust, remains largely unexplored due to the limited research on this topic. Nevertheless, given the connection between trust and delegation [Aggarwal and Mazumdar, 2008; Leana, 1987; Lubars and Tan, 2019; Stout et al., 2014; Sun et al., 2022], factors underlying trust in virtual agents may also influence delegation to virtual agents in similar ways, either directly or mediated through trust. To approach and explore these factors, we proposed a conceptual model of trust in virtual agents and extended it to cover also delegation. The model considers various factors and provides an explanation of how they collectively shape users' trust in and delegation to virtual agents.

Before elaborating on the model and its extension in Section 3.3, this chapter briefly discusses different definitions of trust in Section 3.1 and gives an overview of trust-related factors in Section 3.2.

3.1 Definitions of Trust

The essence of trust has been the subject of much scientific debate. Some observe trust from a sociopsychological perspective, emphasizing its function as a social relation [Luhmann, 1979]. Other researchers, for example, are inclined towards a psychological point of view, stressing the affective basis of trust [Dunn and Schweitzer, 2005; Myers and Tingley, 2016]. There also exist works that model and quantify trust, typically in economics and computer science [Marsh, 1994]. Consequently, the definition of trust diverges across disciplines. Computer scientists, economists, philosophers, psychologists, and sociologists all take their individual perspectives on trust. The remainder of this section lists several frequently cited definitions, each followed by a brief remark.

> The definition of trust [...] is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party. [Mayer et al., 1995]

Mayer et al. argued that trust is an individual's *willingness* to be *vulnerable* to others' actions. This basic idea was later widely accepted in trust research [Cook et al., 2005; Friedman et al., 2000; Rousseau et al., 1998]. As of April 2003, the publication from which this definition originates had already been cited "far more frequently than others on the topic of trust" [Lee and See, 2004]. In November 2023, its number of citations had substantially increased to 31785 according to the statistics from Google Scholar, still outnumbering other commonly mentioned definitions multiple times.

> Trust: confidence that [one] will find what is desired rather than what is feared. [Deutsch, 1973, p. 148]

Deutsch considered trust rooted in the cost-benefit analysis of future events. A rational individual always desires and thus pursues benefits, but fears and thus avoids costs. Trust arises when an individual "is confronted with an ambiguous path, a path that can lead to an event perceived to be beneficial or to an event perceived to be harmful" [Deutsch, 1962, p. 303]. When trusting, an individual holds the prospect that taking this path will be beneficial.

> Interpersonal trust is defined here as 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. [Rotter, 1967]

To Rotter, trust is an expectancy that others will do what they state or promise. He posit that social interaction with others (such as parents, teachers, and peers) provides an individual with rich feedback to validate whether their expectancy is accurate. With feedback accumulating, an individual develops a *generalized expectancy* about the extent to which other people in this world can be relied upon. Some individuals may have been surrounded by genuine persons and believe that people are trustworthy in general. Others may have experienced much dishonesty, which makes them less trusting. This generalized expectancy changes slowly in the long term and comprises a personal trait [Rotter, 1967].

> Trust can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability. [...] an agent can be automation or another person that actively interacts with the environment on behalf of the person. [Lee and See, 2004]

The notion of *trust in automation* has been gaining relevance in trust research as automated systems are extensively used today. There are several definitions of trust in automation. For example, in this definition proposed by Lee and See [2004], the trustee is considered an agent that "can be automation or another person". The use of the term "agent" implies that it is the trustee's agency rather than identity that contributes to the trustor's perception of uncertainty and vulnerability. Thus, with their increasing agency, automated systems can be equally valid trustees as humans.

3.2 Trust in Virtual Agents

As discussed above, trust has been defined in various ways, encompassing willingness, attitude, confidence, expectancy, etc. Nevertheless, the different definitions share a common view that trust is a mental construct, whose formation can be described using an information-processing model [LaViola Jr. et al., 2017, Chapter 3.2], as Figure 3.1 illustrates. The processor, functioning like a black box, takes the information perceived from the external world or introspection as inputs and produces mental constructs as outputs. Following this model, trust in a virtual agent can be viewed as the product of users' perception and process of the information about their interaction with the agent.

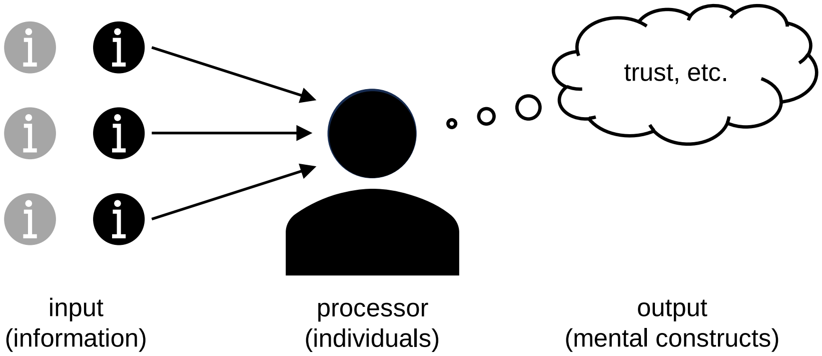


Figure 3.1: Trust in virtual agents as information processing. The gray and black information icons represent information overlooked and perceived by the processor, respectively.

The processor determines how perceived information is processed into trust. Each processor represents an individual user and may respond to perceived information in its unique way. Consequently, the same information may result in different levels of trust depending on the processor, which makes it challenging to predict trust in virtual agents at the individual level. Nevertheless, there are still some general patterns in the trusting attitudes of demographically different groups. For instance, females tend to trust virtual agents more than males [Khalid et al., 2018]. Old people consider a virtual agent more trustworthy than young people when the agent uses non-verbal behavior to communicate its emotions [Hosseinpanah et al., 2018].

On the other hand, the same processor may produce different levels of trust if its perceived information differs. The literature has documented various pieces of information that, when perceived by users, can impact their trust in virtual agents. These pieces of information can be generally classified into two categories –analytical information and affective information– based on an often-used dual division of trust into cognitive and affective aspects [Lewis and Weigert, 1985; Morrow Jr. et al., 2004; Punyatoya, 2019]. *Analytical information* (e.g., the probability of whether an agent can achieve a task) constitutes the basis for users to make rational assumptions about the trustworthiness of a virtual agent. These assumptions give users "good reasons" to trust or distrust another [Morrow Jr. et al., 2004] and are widely regarded as a fundamental component of trust [Lee and Moray, 1992; Mayer et al., 1995; Rempel et al., 1985]. *Affective information* influences trust in virtual agents mainly through psychological and social channels such as emotion, feeling, or intuition. For example, the Uncanny Valley effect can impair virtual agents' trustworthiness by inducing the sense of eeriness [Song and Shin, 2022]. The two categories are not mutually exclusive. Certain information can influence trust both in rational and affective ways. For instance, the urgency of a task not only comprises an important factor to be considered in rational thinking but also puts users under stress that may bias their trust [Monroe and Vangsness, 2022; Potts et al., 2019].

The list below gives some examples of analytical and affective information. Many of them are drawn from research on trust in virtual agents, and the remaining ones originate from studies on trust in automated systems or software agents, to which virtual agents also belong. To facilitate the discussion, within the list, I use the term "agentic system(s)" or simply "system(s)" to generally refer to an artificial entity with some levels of autonomy for carrying out users' tasks. Such an entity can be an automated system, a software agent, or a virtual agent.

Analytical Information Examples

- *Reliability.* The reliability of an agentic system is often demonstrated to be an influential factor in its trustworthiness [Lee and Moray, 1992; Lee and See, 2004; Parasuraman and Riley, 1997; Thielsch et al., 2018]. Users tend to trust reliable systems, whereas signals indicating low-level reliability, such as errors or task failures, are generally detrimental to their trustworthiness [de Visser and Parasuraman, 2011; Manzey et al., 2012]. Notably, users are particularly attentive to mistakes and errors due to the stereotype that machines can perform tasks perfectly [Dzindolet et al., 2002]. Different types of signals indicating low-level reliability may impair trust to varied degrees. For instance, Johnson et al. [2004] found that, in a defect-detection task, trust in an automated system degraded to a larger extent when it reported a false alarm than missing a defect.

- *Predictability.* Predictability describes the extent to which an agentic system's behavior, whether reliable or not, is predictable to its users. Predictability is considered a fundamental factor of interpersonal trust [Mayer et al., 1995] and is also relevant to trust in software agents [Daronnat et al., 2020].

- *Benevolence.* In critical tasks where users' properties or information can be exploited, agentic systems may appear more trustworthy than human agents owing to their user-centered design. For example, Sundar and Kim [2019] found that people are more likely to reveal their credit card information to a software agent than a human agent. Similarly, in massive multi-player online games, players often trade valuable virtual items through a non-player character escrow rather than a third player [Lehdonvirta and Castronova, 2014, p. 122].

- *Transparency.* An agentic system is considered more trustworthy if its algorithms are more transparent, i.e., exposed to rather than hidden from users [Glikson and Woolley, 2020; Hoff and Bashir, 2015]. Transparency is highly relevant to trustworthiness since algorithms has been increasingly complicated. For example, agentic systems ma proactively explain the underlying decision-making process to users to improve the comprehensibility of their algorithms [Dzindolet et al., 2003; Hald et al., 2021].

- *Task difficulty.* Madhavan et al. [2006] found that, in tasks appearing relatively easy, users' trust in and reliance on an automated system may degrade to a larger extent after it fails to accomplish assigned tasks.

Affective Information Examples

- *Anthropomorphism.* Agentic systems are more likely rated trustworthy when they exhibit anthropomorphic features [Natarajan and Gombolay, 2020; Waytz et al., 2014]. For example, virtual agents with humanlike visual representation are perceived as more trustworthy than those embodied in non-human characters [Matsui and Koike, 2021; Sun and Botev, 2021b]. The same holds for the auditory channel: users tend to put a higher level of trust in agents dubbed with natural human voice than low-quality synthesized voice [Chiou et al., 2020]. Other than increasing trust, anthropomorphism can also dampen trust loss after a system fails to accomplish assigned tasks [de Visser et al., 2016].
- *Consistency.* Gong and Nass [2007] found that virtual agents with humanlike visual representation but synthesized voice are considered less trustworthy than virtual agents with humanlike representation and natural human voice or than virtual agents with robotic representation and synthesized voice. Likewise, Nass and Lee [2001] found that voice-based agents are perceived as more trustworthy when their personality cues are consistent with their verbal messages than when they are inconsistent. This consistency effect corroborates research on person perception, where "it was assumed that inconsistency perceived as intrinsic to the character of a person is unpleasant and that such persons will be relatively disliked" [Hendrick, 1972].
- *Similarity.* Individuals tend to hold favorable opinions of those who share similarities with them [Montoya et al., 2008] and, consequently, are more likely to trust those similar others [DeBruine, 2002; Verosky and Todorov, 2010]. This similarity effect exists not only interpersonally but also between users and agentic systems. For example, Verberne et al. [2014; 2015] found that an individual user is more likely to trust a virtual agent if its face resembles the user's face. This effect also applies to behavioral aspects; virtual agents are perceived as more trustworthy when they mimic or synchronize with users' body movements [Launay et al., 2013; Tamborini et al., 2018].
- *Politeness.* When communicating with users, agentic systems that conform to human social etiquette can appear more trustworthy than systems disregarding it. Parasuraman and Miller [2004] found that users are more likely to trust

an automated system that communicates in a relatively non-interruptive manner (e.g., postponing the notification of non-critical messages when users are focusing on important tasks). de Visser et al. [2016] found that apologizing to users can help virtual agents restore their trustworthiness after failing to accomplish assigned tasks. This trust-repairing effect can be stronger if the apology comes with a self-blame for task failures [Buchholz et al., 2017].

- *Attractiveness.* Humans intrinsically tend to label physically attractive individuals with more favorable characteristics, such as being more trustworthy [Patzer, 1983], than average individuals. The same also holds for agentic systems. Yuksel et al. [2017] found that virtual agents with humanlike visual representation are rated more trustworthy if their faces look attractive. According to their experiment results, facial attractiveness is equally or even more influential in trust than reliability.
- *Personableness.* For embodied virtual agents, their trustworthiness increases when they exhibit personable cues, such as an authentic smile [Krumhuber et al., 2007; Luo et al., 2022] or the so-called *smiling voice* [Torre et al., 2020]. However, these cues may also lead to an opposite effect that decreases virtual agents' trustworthiness in serious or high-stake scenarios [Zhou et al., 2019].
- *Other agents.* The trustworthiness of an agentic system can be biased by the presence or information of another system due to the *contrast effect*. Moradinezhad and Solovey [2021] found that, in tasks where users and virtual agents collaborate, a relatively cooperative virtual agent would appear more trustworthy when it was paired with a relatively uncooperative virtual agent than with a similarly cooperative agent. Notably, a software application can be deliberately represented as multiple virtual agents, with each agent carrying a unique component of the application [Baylor, 2009].
- *Eye gaze.* For agentic systems whose visual representation includes eyes, their gazing behavior can influence their trustworthiness. For instance, research shows that people consider a human face more trustworthy if the face more frequently gazes at them than somewhere else [Willis et al., 2011; Wyland and Forgas, 2010]. Similarly, trust in virtual agents increases if they focus on users rather than gazing away [Normoyle et al., 2013].

- *Sympathy.* Lee et al. [2007] found that virtual agents with a caring orientation –e.g., showing sympathy or providing supportive and encouraging feedback to users– are perceived as more trustworthy than those with a neutral orientation.

- *Graphics.* The graphical style of an embodied virtual agent can impact its trustworthiness to some extent [McDonnell et al., 2012; Torre et al., 2019].

3.3 Trust Dimensions

From a macroscopic perspective, there are mainly three dimensions governing virtual agents' trustworthiness in the information-processing model discussed above, including an *analytical dimension*, an *affective dimension*, and a *technological dimension*. The analytical dimension impacts trust in virtual agents via being directly involved in rational thinking, while the affective dimension biases user trust through psychological and social channels such as emotion, feeling, or intuition. During interaction with virtual agents, users play the role of an information processor that continuously perceives and extracts information from the interaction. The perceived information, in turn, shapes their trust in virtual agents via the analytical and affective dimensions, as Figure 3.2 illustrates. Some information exerts its influence mainly through a single dimension, whereas others can have an impact both in the rational and affective aspects.

The technological dimension accounts for the indirect influence of technologies underlying virtual agents. As previously discussed in Section 2, the perception of a virtual agent varies depending on the media device used. According to the information-processing model, different perceptions (i.e., information perceived and extracted from facts) will produce varied levels of trust. Furthermore, technologies underlying virtual agents are highly relevant to data security. Trust in a virtual agent may decrease if the agent runs on an outdated infrastructure threatening personal data and privacy. This issue can be more prominent for immersive media devices whose embedded sensors are usually more invasive than those in laptops or smartphones [Adams et al., 2018].

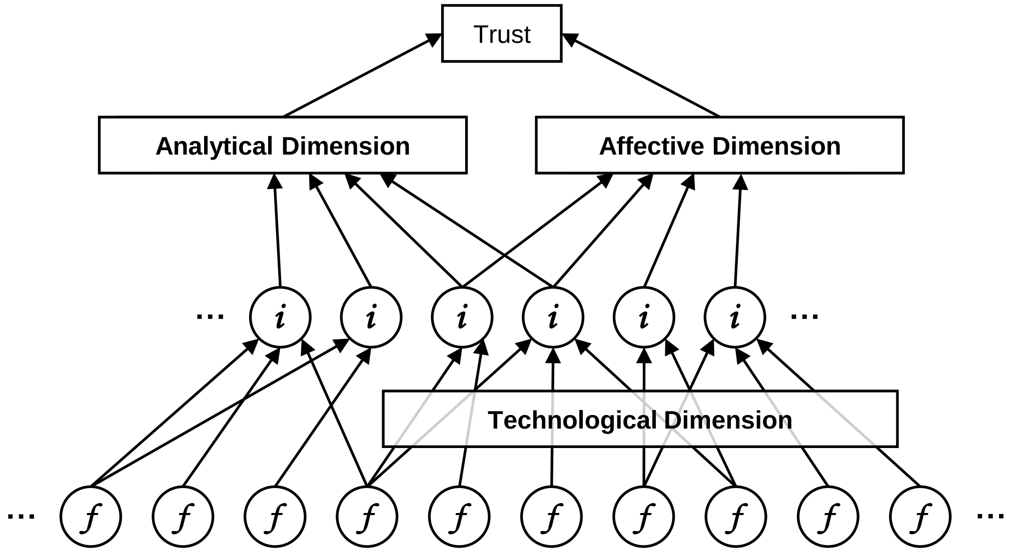


Figure 3.2: A conceptual model of trust in virtual agents. The circled f in the figure bottom represents facts about interaction with virtual agents. These facts are perceived by users and become information (illustrated as circled i). A piece of information may originate from several facts; for example, an agent's capability of financial investment may result from an appraisal of its trading histories in different markets. The bulk of information (represented as the four circled i on the right) is perceived through media devices and thus subject to the technological dimension. The remaining information (represented as the two circled i on the left) is perceived via other channels, such as conversations with other people about the agent's reputation. Perceived information, in turn, influences trust in virtual agents through the analytical and affective dimensions.

Model Extension

Trust and delegation are closely related concepts. Theoretically, delegation entails two essential components of trust that Mayer et al. [1995] put forward: (1) *uncertainty*: agents' actions and the ensuing outcomes are neither entirely predictable nor completely unknown; (2) *vulnerability*: users are accountable or responsible for task outcomes.

Other than their common theoretical basis, there is also empirical evidence substantiating the connection. For example, trust was found correlated with the level of *decentralization*, i.e., the extent to which top managers of an organization empower subordinates to make decisions as opposed to micro-managing. Research shows that the more trust there is within an organization, the more decentralized the organization is, and the more delegation therein [Gur and Bjørnskov, 2017]. As noted in [Fukuyama, 1996, p. 31], "a high-trust society can organize its workplace on a more flexible and group-oriented basis, with more responsibility delegated to lower levels of the organization. Low-trust societies, by contrast, must fence in and isolate their workers with a series of bureaucratic rules." Furthermore, taking a more psychological perspective, Leana [1987] and Aggarwal and Mazumdar [2008] found that trust in subordinates is a vital factor determining managers' decisions on delegation. In the context of human-computer interaction, trust also plays an important role in delegation to artificial agents. Stout et al. [2014] found that trust is correlated with students' willingness to let a software agent make the travel arrangement for a job interview. Lubars and Tan [2019] identified trust as a basic constituent of delegability (cf. Section 1.2). The trust-delegation correlation was also found in our research (cf. Chapter 6).

Given the evidence mentioned above, the three dimensions may also influence delegation to virtual agents similarly to how they impact trust in virtual agents. Thus, the conceptual model can be further extended to cover delegation, as Figure 3.3 illustrates, where the three dimensions influence delegation to virtual agents directly or through the mediation of trust. The extended model provides a theoretical foundation to systematically view and explore factors potentially governing delegation to virtual agents.

Limitations

In Figure 3.2, the cognitive process of information perception is illustrated as arrowed lines between facts and information, whereas the actual relationship is much more complicated than a linear one [Weimann, 1999, p. 62]. Furthermore, Figure 3.2 implies that each piece of perceived information will eventually influence trust in one way or another. However, in reality, people do not traverse their memory and use all the available information to make judgments; instead, they tend to utilize the most accessible or focused bulk [ibid., p. 65].

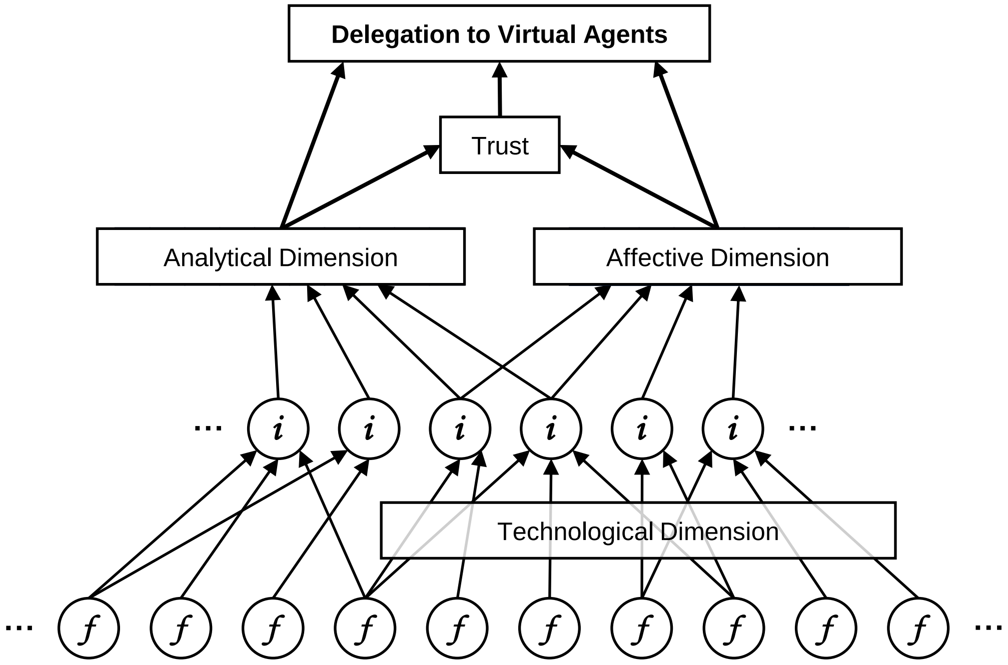


Figure 3.3: The conceptual model extended from trust to delegation.

3.4 Chapter Summary

This chapter presented an overview of trust-related literature and elaborated on a conceptual model, where trust in virtual agents is described as a process of users perceiving and processing information. The model distinguishes three dimensions, each representing a relatively unique aspect of trust in virtual agents. Other than the often discussed cognitive and affective dimensions, the model also includes a technological dimension to acknowledge and highlight the potential influence of rapidly developing media technologies on virtual agents' trustworthiness. The model is further extended to provide a theoretical foundation for systematically exploring factors of delegation to virtual agents.

With the theoretical foundation and various potential factors laid out in this chapter, the next chapter investigates which ones are the most relevant to users' decisions on delegation to virtual agents.

Chapter 4

Exploring Relevant Factors of Delegation to Virtual Agents

Portions of the content presented in this chapter were published in [Sun and Botev, 2021c].

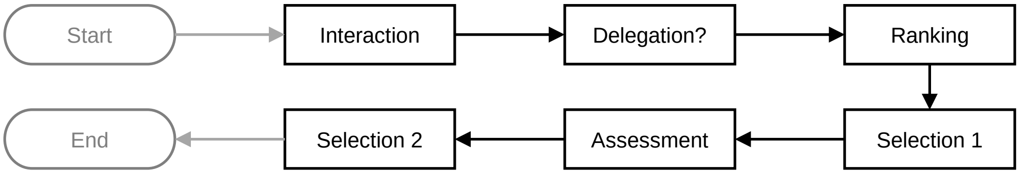
The previous chapter discussed numerous relevant factors, each potentially having its unique impact on delegation decisions to virtual agents. Knowing which ones are more influential than others is informative and can facilitate the study and design of virtual agents. This chapter presents an exploratory study that aims to reveal the most relevant factors in delegation to virtual agents.

4.1 Methodology

Eighty native English speakers were recruited on Prolific[1] and asked to complete a user study that consisted of two parts (cf. Figure 4.1). In the first part, participants briefly interacted with a virtual agent, during which they faced a decision on whether to delegate a critical task to the agent. The interaction was experienced in a 3D virtual environment (cf. Figure 4.2a) on laptops or desktop computers. The environment was minimally decorated to prevent participants from being distracted by the surroundings. The agent was animated using motion capture

---

[1] As of 2021 when the study was conducted, Prolific was a crowd-sourcing platform dedicated to scientific experiments.



#	Step	Detail
1	Interaction	Participants interact with the agent.
2	Delegation?	Participants decide whether to delegate.
3	Ranking	Participants rank the 20 factors individually.
4	Selection 1	Participants select major and minor factors among rank-A and rank-C factors.
5	Assessment	Participants assess the agent regarding each of the rank-A and rank-C factors.
6	Selection 2	Participants select major and minor factors among rank-A and rank-C factors again.

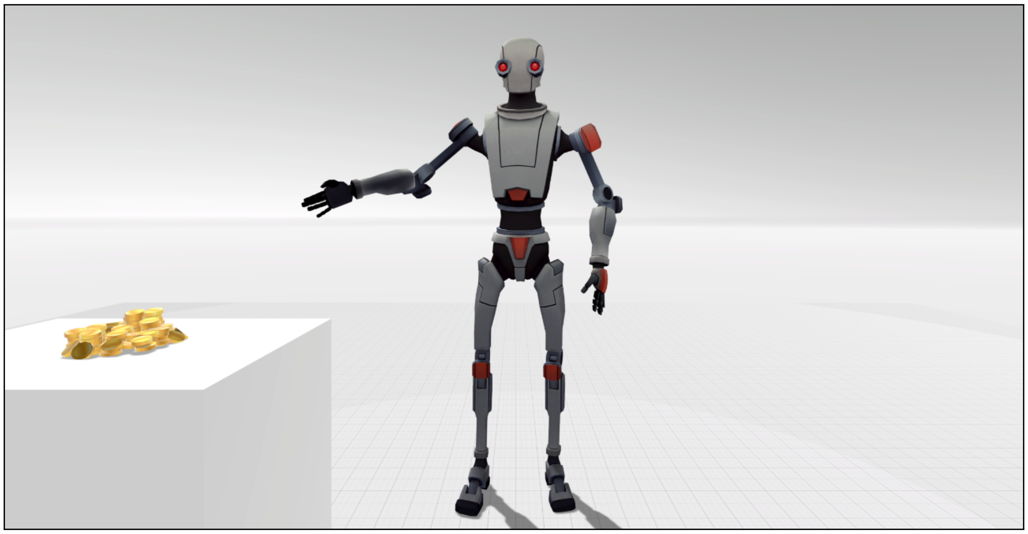
Figure 4.1: Flowchart showing the user study procedure.

to make the interaction feel natural. Participants' position in the environment was fixed, and their view was oriented toward the agent to simulate face-to-face conversations. In the second part of the user study, participants answered a survey that probed into their delegatory decisions made during the interaction.

Participants were initially endowed with some virtual gold coins, which were placed on a virtual table next to the agent (cf. Figure 4.2a). The coins were framed as tokens to exchange for a completion code at the end of the survey. The completion code, in turn, could be exchanged for a monetary reward on Prolific.

During the interaction, the agent was embodied in a virtual robot and spoke to participants using a synthesized male voice. Its speech started with a brief self-introduction and then to an explanation of the coins' usage, following which the agent made a proposal: if participants gave the agent all the coins they had, the agent would return them a different completion code doubling the monetary reward. However, the agent also warned participants about the potential risk of failure, in which case it would return a completion code worth only half the reward. This dynamic remuneration was fabricated to simulate a critical context similar to investing money in or through others. Participants were treated equally with the standard amount of money.

After the decision was made, participants were presented with a survey and some instructions. At first, 20 words (cf. the "Factor" column in Table 4.1)



(a) Interacting with the agent.

What do you think of the agent's adaptivity? Use the slider below to let us know what you think.

(b) Assessing the agent regarding its adaptivity.

Figure 4.2: First-person perspective screenshots of the virtual environment.

were displayed to participants in sequence. Each word represents a unique factor potentially related to users' delegation decisions. Participants were asked to recall their decision-making process and rank each factor according to the following rules: (1) label a factor with *rank A* if the factor was consciously, clearly, or explicitly considered when deciding; (2) label a factor with *rank B* if the factor never came to mind when deciding; (3) label a factor with *rank C* if the factor does not fit any of the two ranks above.

Table 4.1: Factors involved in the survey. The references in the table provide evidence that these factors are related to delegation decisions.

Factor	**Opposing Poles**	**Reference**
Adaptivity	adaptive – non-adaptive	[de Visser et al., 2011]
Aliveness	animate – inanimate	[Sun and Botev, 2021b]
Attractiveness	attractive – unattractive	[Yuksel et al., 2017]
Autonomy	autonomous – non-autonomous	[Verberne et al., 2012]
Capability	capable – incapable	[Leana, 1987]
Compliance	compliant – non-compliant	[Eisenhardt, 1989]
Confidence	confident – unconfident	[Sundström et al., 1994]
Controllability	controllable – uncontrollable	[Stout et al., 2014]
Eeriness	eerie – reassuring	[Mori, 1970]
Human-likeness	humanlike – non-humanlike	[Straßmann et al., 2018]
Informativeness	informative – uninformative	[Verberne et al., 2012]
Informedness	informed – uninformed	[Hoff and Bashir, 2015]
Moral quality	benevolent – malevolent	[Mayer et al., 1995]
Observability	observable – unobservable	[Eisenhardt, 1989]
Politeness	polite – impolite	[Parasuraman et al., 2004]
Predictability	predictable – unpredictable	[Muir, 1994]
Reliability	reliable – unreliable	[Yuksel et al., 2017]
Reputation	reputable – disreputable	[Hoff and Bashir, 2015]
Trustworthiness	trustworthy – untrustworthy	[Stout et al., 2014]
Usability	easy-to-use – difficult-to-use	[Hoff and Bashir, 2015]

The ranking process serves to filter out overlooked factors (i.e., rank B) in participants' decision-making process. The remaining ones (i.e., rank A and C) might have impacted participants' delegation decisions to different degrees. To examine which ones were more relevant, participants were asked to pick out *major factors* and *minor factors* that largely and minimally impacted their decisions, respectively. Afterward, participants assessed the agent regarding each of the remaining factors. For example, participants might need to specify how much they think the agent was reliable. The assessment was measured on a visual analog scale, as Figure 4.2b depicts. The two endpoints of the sliding bar denote the two poles of a factor (cf. Table 4.1). Participants could pick a point on the bar to indicate their assessment of the factor relative to the two poles. This assessing process forced participants to reflect on the remaining factors. Once the assessment finished, participants repeated the process of selecting major and minor factors. The difference between before- and after-assessment selection results would reveal whether a short period of reflection could change participants' perceptions of these factors' impacts.

4.2 Results

Participants ($N = 80$) are diverse in demographics, such as age (mean = 33.3, SD = 11.2), gender (34 males, 46 females), and country of residence (mainly Canada, the United Kingdom, and the United States). Most participants (70 out of 80) chose to give the coins to the agent.

Figure 4.3 illustrates the proportions of the different ranks that each participant gave in the ranking process. There were evident individual differences in the proportions of rank-A and rank-B factors, whereas the proportions of rank-C factors were generally small and relatively consistent across participants. Participants on average consciously or explicitly considered 9.0 factors (SD = 3.63; Shapiro-Wilk test: $W = 0.974, p = 0.098$). The number of overlooked factors per participant followed a similar distribution with a mean value of 7.7 (SD = 3.79; $W = 0.959, p = 0.012$). Only 3.3 factors (SD = 2.54; $W = 0.926, p < 0.001$) were in the limbo state between being overlooked and consciously considered.

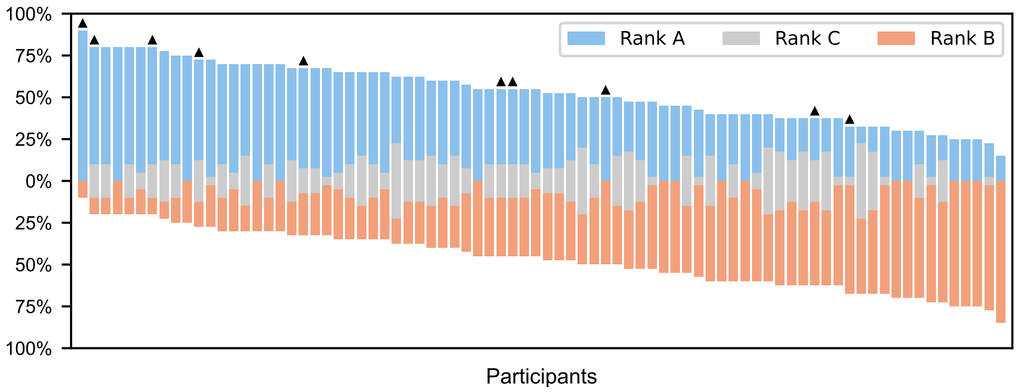


Figure 4.3: Ranking proportions per participant. Each bar represents a unique participant. Those who refused to delegate are marked with a black triangle on top of the bar. Each bar is composed of maximally three colors showing the proportions of rank-A (consciously considered), rank-B (overlooked), and rank-C (neither overlooked nor consciously considered) factors.

Factors frequently labeled with rank A include informativeness, capability, informedness, politeness, reliability, and trustworthiness (cf. Figure 4.4). Among them, informativeness was the most relevant factor as it was explicitly considered by 75% of participants, a proportion fairly higher than that of others, such as capability (66%), politeness (58%), and trustworthiness (57%). Its prominence was also reflected in the process of selecting major and minor factors (cf. Figure 4.5), where informativeness received the highest number of votes as a major factor both before and after the assessment. These results suggested that virtual agents' informativeness is important in delegation decisions to virtual agents. Its influence may even surpass other acknowledged key factors early during the interaction. Additionally, other common rank-A factors (e.g., politeness, reliability, and trustworthiness) were also frequently selected as major factors.

On the other hand, often overlooked factors include moral quality (overlooked by 51% of participants), attractiveness (48%), adaptivity (46%), reputation (46%), eeriness (46%), aliveness (46%), and confidence (45%). Although these factors were overlooked by approximately half of the participants, there were still many who either consciously considered them or were somewhat influenced by them.

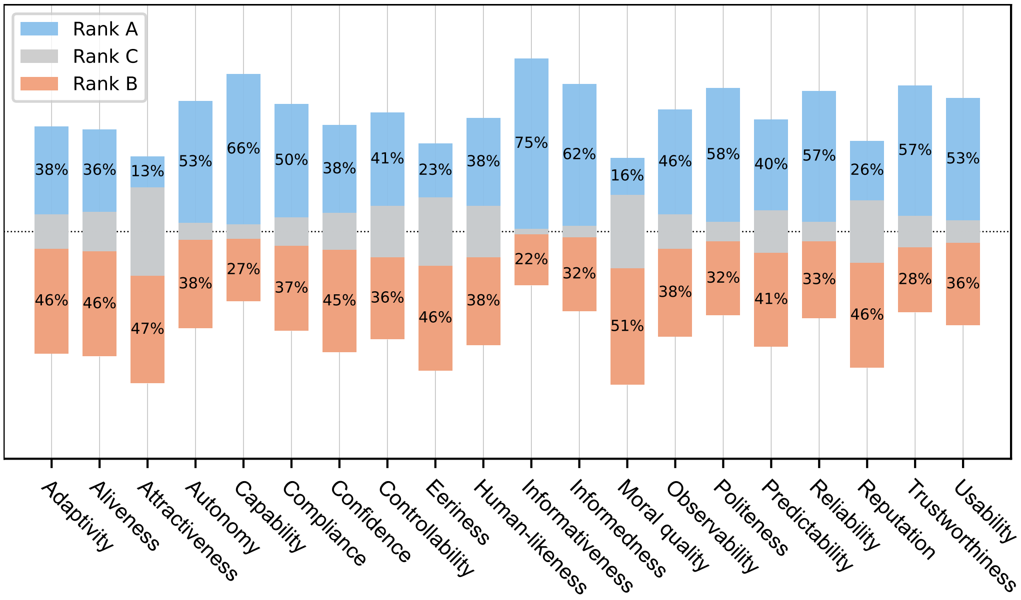


Figure 4.4: Ranking proportions aggregated by factor. Each bar represents a unique factor and its aggregated ranking results over all participants.

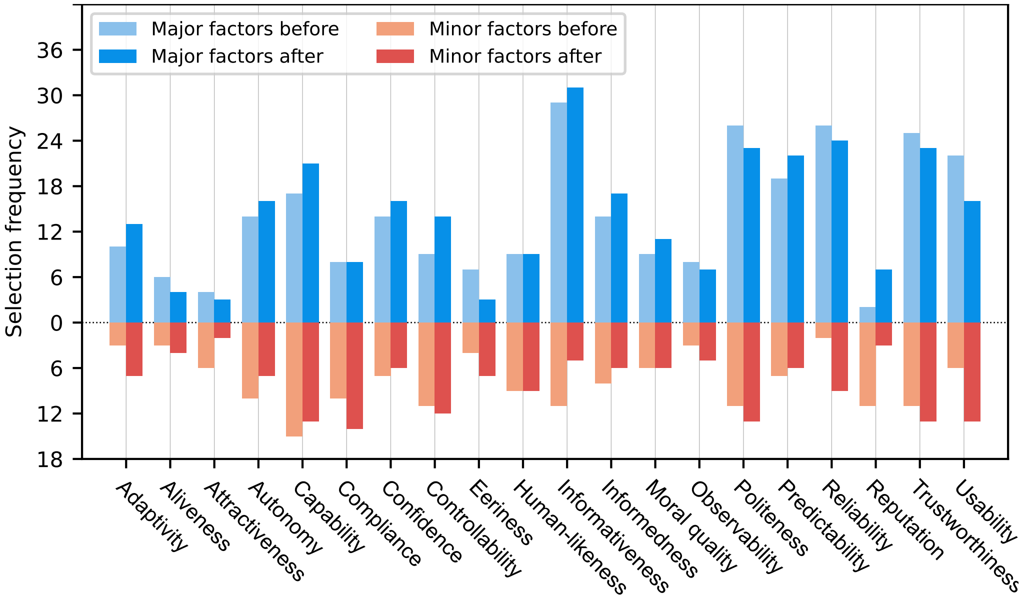


Figure 4.5: Selection frequency (major/minor, pre/post assessment).

As Figure 4.5 indicates, the assessment process did not significantly change participants' selection of major and minor factors, except that usability was more likely to be a minor factor in the post-assessment selection (Wilcoxon signed-rank test: $W = 10, p = 0.010$). There was also a notable difference in the votes for reputation before and after the assessment, but the sample size was small.

4.3 Discussion

The result that most participants chose to delegate contradicts the previous finding that people prefer to retain control and are reluctant to delegate (cf. Section 1.1). The predominant preference for delegation can be attributed to the small amount of monetary reward. Halving or doubling it makes little difference to the final remuneration, especially considering that most participants live in developed countries. The relatively non-critical consequences might have encouraged participants to delegate and take the chance.

Although most participants made the same decision on delegation, their decisions were made at varied levels of awareness, as manifested in the considerable individual differences in the number of rank-A factors per participant. Such differences may originate from participants' heterogeneous dispositions of being thoughtful or different understandings of the term "consciously considering" (from momentary thoughts to intensive reasoning).

Most participants considered the agent's informativeness an essential factor insofar as it outweighed other key factors of delegation, such as capability. The focus on informativeness might emanate from participants' intention to learn about the agent early during the interaction. However, since participants only briefly interacted with the agent, it remains unclear whether the influence of informativeness would persist if the interaction proceeded. Its influence is arguably at its peak early during the interaction and will decrease as users become more familiarized with virtual agents.

Among the overlooked factors, those related to the agent's visual properties (attractiveness, eeriness, and aliveness) might have biased participants' decision-making process subconsciously rather than being directly involved in their reasoning, which may explain why many participants did not consciously relate these factors to their delegation decisions. The less consideration of moral quality was likely due to the common belief that machines are user-centered. Participants were only provided with scant information about the agent's background, which might have caused the low-level awareness of reputation.

Implication for the Conceptual Model (cf. Chapter 3)

Informativeness is highly relevant to the rational aspect. Its prominence indicates the impact of the analytical dimension on delegation.

Limitations

To prevent participants from being bored or frustrated by a lengthy survey, each factor was assessed only using a single item. This approach provided a quick estimation of participants' factor assessments, however, at the cost of reduced reliability. Moreover, the survey did not include any item for filtering out inattentive participants and speed runners, which further lowered the data quality. To yield more reliable results, one may consider using more fine-grained measures (e.g., using or adapting established questionnaires), adding attention checks, and focusing on fewer factors (e.g., 10 instead of 20).

The survey investigated a specific selection of 20 agent-related factors, and the ranking results were completely based on participants' subjective responses, i.e., their consciousness. This setting omitted the influence of subconsciousness and other factors related to dispositions or environments, where more influential factors may exist. These aspects should also be considered, for example, by using behavioral and physiological measures to detect subconscious effects or by replacing some factors in the survey with dispositional or environmental ones (e.g., extroversion or task difficulty).

The interaction context used in the survey was a simple, abstract representation of many real-world scenarios involving delegation. While it is adequate for scientific investigation, delegation decisions in reality are usually made in more complex situations. Thus, findings derived from this survey may only have a low level of ecological validity. More evidence based on experimental or field studies is needed to validate these findings. For example, it will be highly informative to conduct an experiment that manipulates a virtual agent's informativeness and tests its effect on delegation to the agent.

The interaction between participants and the agent was short, due to which participants might find it difficult to assess some factors, such as reputation and predictability. These concepts are usually formed over time. Follow-up research may consider prolonging the interaction sufficiently to capture the full dynamics of these factors.

4.4 Chapter Summary

This chapter presented a survey-based study exploring which factors are the most relevant to users' decisions on delegating to virtual agents. A selection of 20 factors was tested, among which informativeness, politeness, capability, reliability, and trustworthiness were frequently reported to be explicitly or consciously considered in participants' decision-making process. The results suggest that informativeness is the most salient factor early during the interaction, insofar as its impact may surpass those of other acknowledged factors, such as capability and trustworthiness. Thus, for virtual agents carrying out critical tasks, developers may consider letting the agents supply sufficient information to users.

The next chapter presents another survey-based study investigating the effect of an individual factor –virtual agents' visual representation– on delegation.

Chapter 5

The Impact of Visual Representation on Delegation to Virtual Agents

Portions of the content presented in this chapter were published in [Sun and Botev, 2021b].

As previously discussed in Section 2.1, the visual representation of a virtual agent can take countless forms, each potentially having its unique impact on delegation decisions. Knowing these impacts is essential because virtual agents are often embodied. However, little research has been dedicated to revealing them. This chapter elaborates on a survey-based study investigating user preference for the visual representation of virtual agents that carry out critical tasks on users' behalf. Several commonly used categories of representation were differentiated and compared.

5.1 Methodology

A survey was conducted through Amazon Mechanical Turk. Initially, participants were given a short brief –*imagine that you have a large amount of money to invest through a virtual agent*– to simulate a critical context of delegating financial investment to a virtual agent. After reading the brief, participants were presented with five categories of agent representation (cf. Figure 5.1) distinguished based on their visual resemblance to humans. In a descending order of human-likeness, the five categories were human, humanoid, organic non-human, inanimate object, and symbol. Participants were asked to give each category a unique number from one to five, where number one denotes the most favorite agent representation to interact with for the investment, and number five denotes the least favorite.

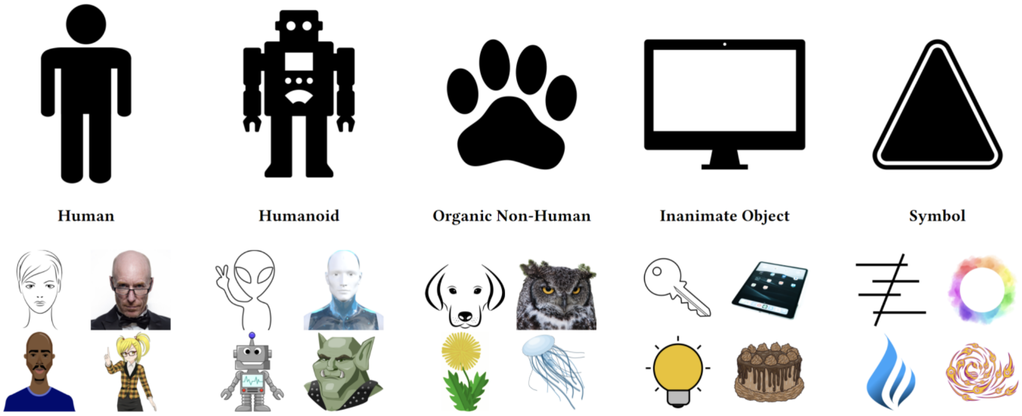


Figure 5.1: The five categories of agent representation used in the survey. From left to right, their human semblance decreases. Above each category's name is an illustration of the category. Under the name are four images showing examples of different visual fidelities.

For each category, four specific examples were provided to demonstrate the varieties therein. In the human category, the four examples varied in their age, gender, race, haircuts, dressing, and more. Besides, the examples contained different visual fidelities, including line drawing (the top-left example), photorealistic image (top-right), symbolic low-detail drawing (bottom-left), and descriptive high-detail drawing (bottom-right). Through these examples, the survey involved a broad spectrum of visual representations, covering not only those frequently investigated in research (e.g., human, robot, animal, and object; see [Gulz and Haake, 2006; Straßmann and Krämer, 2017, 2018]) but also two often neglected ones: symbol and non-zoomorphic species (e.g., plant).

The categories and examples were presented to participants in the same layout as the one depicted in Figure 5.1, except that the order of the categories was randomized for each participant. For instance, one participant might see the sequence “human, humanoid, organic non-human, inanimate object, and symbol”, while another participant might encounter the sequence “symbol, organic non-human, human, humanoid, and inanimate object”. The randomization was to prevent the category order from biasing the results.

5.2 Results

Two groups of participants were recruited to compare user preference for agent representation in different demographics: a *savvy group* who had purchased video games before and a *general group* without any restriction on their backgrounds. To collect high-quality data on Amazon Mechanical Turk, the recruitment was only open to Master Workers, i.e., people who have successfully and diligently performed a large number of tasks on the platform. However, as of the time when the survey was conducted, Master Workers were sparse, and the additional video-game filter made it particularly difficult to gather Master Workers for the savvy group. This led to lopsided sample sizes: 85 participants in the general group whereas only 15 in the savvy group.

Table 5.1 gives the number of the different ranks each category received, and Figure 5.2 presents the data more intuitively in a diverging bar chart. From the most to the least preferred representation, the ranking results in the general group were human, humanoid, inanimate object, symbol, and organic non-human. The results in the savvy group were the same, except that the symbolic and organic non-human categories switched their positions in the sequence.

When the two groups’ data were aggregated, a dividing line was observed between the preference for humanlike categories (human and humanoid) and non-human categories (organic non-human, inanimate object, and symbol). The humanlike categories received mostly positive rankings (1 and 2, i.e., the most and second most preferred), whereas the non-human categories were largely labeled with neutral (3) or negative rankings (4 and 5, i.e., the second least and least preferred). This observation was further supported by a Friedman test on the aggregated data ($\chi_F^2(4) = 30.82, p < 0.01$) with Nemenyi test as the post-hoc analysis (cf. Table 5.2). The ranking differences between the humanlike and non-human categories were mostly statistically significant, except that there was no significant difference between the humanoid and inanimate object representations ($p = 0.136$). Within the humanlike or non-human categories, the ranking differences were not statistically significant.

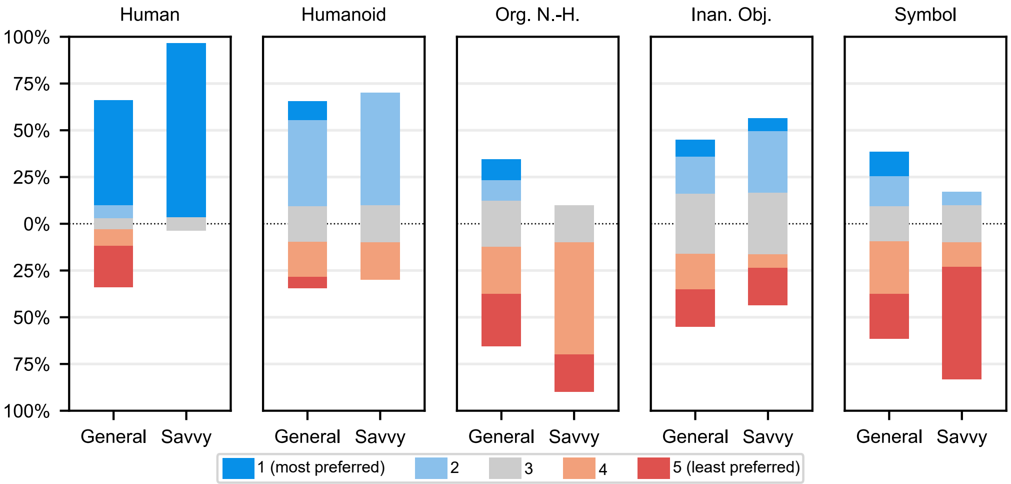


Figure 5.2: Ranking proportions per category.

5.3 Discussion

The survey results indicate that the visual representation of virtual agents can influence delegation decisions. Users generally prefer to interact with humanlike virtual agents, such as those embodied in human or humanoid characters. The willingness to interact declines substantially when the visual representation has a low level of anthropomorphism. The willingness can further decrease if zoomorphic features are present. These effects were found not only in general users but also in those who are more familiar with virtual environments. However, given the lopsided sample sizes, further investigation is needed to validate whether these effects are cross-demographics.

The preference for humanlike over non-human representations can be explained by the *similarity effect*, which states that humans inherently tend to

Table 5.1: Ranking distributions.

Category	**Group**	**Rank**				
		1	2	3	4	5
Human	General	47	6	5	8	19
	Savvy	14	0	1	0	0
Humanoid	General	9	39	16	16	5
	Savvy	0	9	3	3	0
Org. N.-H.	General	10	9	21	21	24
	Savvy	0	0	3	9	3
Inan. Obj.	General	8	17	27	16	17
	Savvy	1	5	5	1	3
Symbol	General	11	14	16	24	20
	Savvy	0	1	3	2	9

Table 5.2: Post-hoc test (Nemenyi) p-values.

	Human	Humanoid	Org. N.-H.	Inan. Obj.	Symbol
Human	1.000	0.772	0.001	0.005	0.001
Humanoid	–	1.000	0.005	0.136	0.034
Org. N.-H.	–	–	1.000	0.772	0.900
Inan. Obj.	–	–	–	1.000	0.900
Symbol	–	–	–	–	1.000

positively evaluate entities similar to them [Montoya et al., 2008]. Human and humanoid representations visually resemble humans, which might have made participants prefer them over non-human representations for delegation. Notably, although organic non-human representation is visually and biologically more relevant to humans than objects and symbols are, it was considered the least preferred option in the survey. This was also reported in [Löffler et al., 2020], where the likability of animal-like social robots is significantly higher when their visual animal-likeness is around 20% than when it is around 90%.

The aversion to zoomorphic features may emanate from humans' intrinsic vigilance in wild animals, and evidence shows that people tend to treat virtual animals similarly to how they would treat real animals [Melson et al., 2005].

Implication for the Conceptual Model (cf. Chapter 3)

Visual representation is mainly related to the affective aspect. The dividing line between users' preference for humanlike and non-human representations suggests that the affective dimension plays a role in delegation decisions.

Limitations

A major and evident limitation is the highly asymmetric group sample sizes, due to which findings derived from the group comparison are only indicative and need to be validated fully. Since recruiting savvy users on Amazon Mechanical Turk was difficult, one may consider using targeted advertising, such as distributing the survey in social media groups related to video games. Alternatively, one may change the criteria for screening savvy users to, for example, how many hours spent on video games weekly.

The representation categorization used in the survey constitutes a potential limitation as it may not unequivocally capture certain representation types. For instance, an anthropomorphic object is difficult to classify into any of the five categories but falls somewhere between a humanoid and an inanimate object. Future studies may consider employing a more fine-grained classification, for example, by further dividing the humanoid category into more specific ones such as robot, mystical creature, and anthropomorphic object.

Besides, the categories were distinguished based on their human semblance. This semblance contains many aspects, such as the similarity in visual anthropomorphism, perceived consciousness, intelligence, and more. The survey only focused on visual anthropomorphism in general and thus was unable to reveal whether and to which extent other aspects account for participants' preferences. More empirical studies employing different categorizing criteria are needed to answer these questions.

The visual representations used in the survey were static, i.e., not animated, yet the impact of the dynamic aspect is non-negligible. The same animation may deliver different messages when applied from one agent representation to another (e.g., a crying infant vs. a crying adult). Moreover, in reality, most virtual agents are animated rather than staying static, even if the animation is severely limited to a few frames. Thus, the dynamic aspect should be incorporated in the survey, for example, by adding a smiling animation to all the representations involved.

5.4 Chapter Summary

This chapter presented a survey-based study investigating user preference for the visual representation of virtual agents that carry out critical tasks on users' behalf. The survey tested five categories of representation, including human, humanoid, organic non-human, inanimate object, and symbol. The results indicate that users prefer delegating to virtual agents embodied in humanlike representations (human and humanoid) over non-human representations (organic non-human, inanimate object, and symbol). Thus, developers may consider adopting anthropomorphic features in agent representation and avoid zoomorphic elements, which may induce adverse effects.

Following this chapter on the static and visual aspects of virtual agents, the next chapter moves onto a more dynamic and behavioral realm concerning their capacity to use theory of mind.

Chapter 6

Theory of Mind and Delegation to Virtual Agents

Portions of the content presented in this chapter were published in [Sun et al., 2022].

Imagine that a person is presenting at a computer science conference. The slides are projected onto a big screen from a laptop. Suddenly, the laptop stops projection despite several attempts to reconnect. While waiting for the conference volunteer to fix the problem, the person may try to ease the awkward situation by saying: "my laptop cannot stand my tedious presentation and wants me to shut up." The person certainly knows that the glitch was caused by technical issues instead of the laptop being irritated by the presentation. In saying so, the person is using mentalistic terms (e.g., intention, thought, and other cognitive or affective states) to explain the laptop's behavior. This ability to attribute mental states to others is often called *theory of mind* (ToM).

The term ToM was initially coined by Premack and Woodruff [1978] to study chimpanzees' ability to understand the intentions of others. Later, the term was dispersed to interpersonal contexts and now commonly refers to an individual's ability to represent other individuals using mental states (e.g., belief, intention) and distinguish them from one's own mental states. It is a "theory" of mind because others' minds are not directly observable and can only be inferred from observable cues, such as behavior and utterance. Also, mind incorporates a complex of elements, including belief, emotion, thought, intention, etc. To represent such a complicated system is essentially to construct a theory about it.

For most people, ToM develops naturally in early childhood [Russell et al., 1998] and remains an important ability for social interaction [Lecce et al., 2017; Watson et al., 1999]. Many social behaviors intrinsically entail ToM [Perez-Osorio et al., 2021], such as deception or persuasion where knowing and predicting others' mental states are vital. Due to its importance for interpersonal relationships, many researchers are attempting to integrate ToM into artificial agents (e.g., virtual agents, robots) to improve user-agent interaction. Evidence shows that artificial agents with a ToM are perceived as more humanlike [Yoshida et al., 2008], trustworthy [Ruocco et al., 2021], and socially interactive [Perez-Osorio et al., 2021; Sturgeon et al., 2019] than those without a ToM. These effects point to the possibility that a virtual agent's ToM may impact users' decisions on delegating to the agent, yet whether this impact exists is still unclear.

This chapter presents an experimental study that aims to answer this question. The experiment tests a hypothesis that *users are more likely to delegate critical tasks to virtual agents with a high-level ToM*. Before elaborating on the experiment, this chapter introduces more background on ToM in Section 6.1.

6.1 Related Work on Theory of Mind

When discussing ToM, an often mentioned subject is *Sally-Anne test*, a widely used psychological test to indicate an individual's ability to use ToM [Baron-Cohen et al., 1985]. In this test, a human subject is described a story where its two protagonists, Sally and Anne, were in the same room. In front of them were two baskets and a marble. Sally placed the marble in one of the baskets and then left the room. Before Sally came back, Anne moved the marble to the other basket. At this point, the subject is asked "when Sally came back, which basket would she look for the marble?" The correct answer is the same basket that Sally initially placed the marble in because she was away and therefore did not know that the marble had been moved. However, toddlers often fail the test due to their still-developing ToM, which leads them to believe that Sally had the same belief as theirs [Wellman et al., 2001]. Apart from Sally-Anne test, there are also many other tests or tasks for examining whether and to which extent an individual has a ToM (see [Quesque and Rossetti, 2020] for a comprehensive summary).

ToM can be classified into different orders, of which common ones include zero, first, and second orders (cf. Figure 6.1). An individual with a zero-order ToM is unaware of and unable to infer others' mental states. An individual with a first-order ToM can infer others' mental states and distinguish them from the individual's own. An individual with a second-order ToM can infer not only another person's mental states but also this person's mental representation of others. The order of a ToM is theoretically infinite. However, for human brains, using a high-order ToM requires a substantial amount of cognitive effort, which makes it increasingly unfeasible as the ToM order increases.

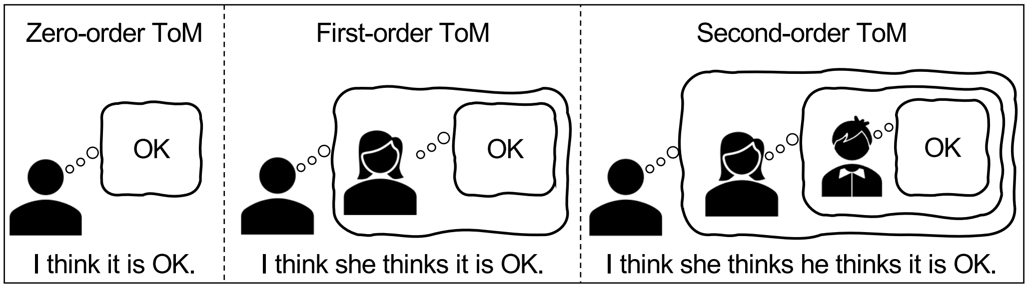


Figure 6.1: Different ToM orders.

ToM and Virtual Agents

ToM is an umbrella term describing the ability to represent others using mental states such as belief, intention, emotion, and more. There are numerous studies trying to enable virtual agents with a specific aspect of ToM, yet many of these studies were not conducted explicitly in the name of ToM. A representative research line of such is *affective computing*, which strives to make machines capable of inferring users' affective states based on observable cues, including facial expression [Tarnowski et al., 2017], body language [Santhoshkumar and Geetha, 2019], eye movement [Lim et al., 2020], EEG signal [Pei and Li, 2021], and more. ToM elements can also be seen in virtual agents for autonomous driving (e.g., lane-changing intention detection [Morris et al., 2011]), healthcare (e.g., suicidal behavior detection [Martínez-Miranda et al., 2019]), video games (e.g., player intention detection [Doirado and Martinho, 2010]), and more.

Evidence shows that virtual agents can benefit from having a ToM. To give a few examples, Narant et al. [2019] found that users felt more comfortable when interacting with virtual agents having a ToM than those without a ToM. Tanaka et al. [2023] found that virtual agents with a ToM were assessed more positively in terms of their trustworthiness and acceptability. Hoogendoorn and Soumokil [2010] found that, in video games, interaction with non-player characters was rated more engaging if they had a ToM.

ToM and Trust

Although there is limited research on the relationship between ToM and delegation, the literature shows that ToM is relevant to trust, a factor closely related to delegation. According to the widely cited trust definition proposed by Mayer et al. [1995], there are at least two fundamental abilities underpinning trust, including the ability to estimate risks and the ability to predict others' actions. The latter involves the inference of belief and intention, which is essentially ToM. Apart from their theoretical connection, there also exists empirical evidence substantiating the ToM-trust relationship. For example, children's ToM is related to their performance in *selective trust* tasks [DiYanni et al., 2012; Rotenberg et al., 2015]. Children with a developed ToM tend to be more reflective and cautious about their trust in others, which leads to a relatively low trusting tendency [Di Dio et al., 2020]. The ToM-trust relationship is also supported by neuroscientific research. Prochazkova et al. [2018] found that an individual's trust can be modulated by stimulating a specific set of brain regions, namely the *theory of mind network*, including precuneus, temporo-parietal junction, superior temporal sulcus, and medial prefrontal cortex. Through fMRI, these regions were found to underlie one's ability to distinguish false beliefs [Saxe, 2009].

While the trustor's ToM is influential in a trusting relationship, the effect of the trustee's ToM remains unclear. Having a ToM may increase a virtual agent's trustworthiness by facilitating the development of a *shared mental model* with human users. A shared mental model can be generally described as a belief that all the members of a team hold in common. This belief can be task-related (e.g., task requirements or difficulties) or about the team per se (e.g., teammates' backgrounds, intentions, or preferences). In a hybrid team consisting of humans and virtual agents, the establishment of a shared mental model can significantly increase humans' trust in virtual agents [Hanna and Richards, 2018]. Virtual agents with a ToM have a higher capacity to build a shared mental model with users by modeling their intentions, beliefs, and more.

A recent study found that users tend to send more money to a virtual agent in the *investment game* when the agent appears to have a ToM than when not [Ruocco et al., 2021]. However, user trust may also decrease when virtual agents possess a high-order ToM, similar to the Uncanny Valley issue [Stein and Ohler, 2017].

6.2 Methodology

A between-subjects experiment was conducted through Prolific to validate the hypothesis that users are more likely to delegate critical tasks to virtual agents with a high-level ToM. Participants played a game, namely *colored trails*, in a 3D virtual environment experienced on laptops or desktop computers.

6.2.1 Colored Trails

The colored trails game was initially proposed in [Grosz et al., 2004] and later formalized in [Gal et al., 2005], serving as a testbed to "design, learn, and evaluate players' decision-making behavior as well as group dynamics in settings of varying complexity" [ibid.]. The game has several variants. This experiment followed the game setting used in [de Weerd et al., 2017], where two players play on a game board consisting of 25 colored tiles, as Figure 6.2 illustrates. Each tile was painted with one of the following five colors: yellow, purple, gray, black, and white. Each player had a game piece initially located on the tile at the board center and was tasked to move the game piece to a unique goal location unbeknownst to the other player. A game piece could be moved one step at a time from its current location onto a horizontally or vertically adjacent tile if and only if the player spent a chip of the same color as the tile. Both players were endowed with four chips at the beginning of the game. To encourage negotiation (cf. the subsequent paragraph), the colors of these chips were manipulated to prevent players from reaching their goal locations only with the initial chip sets.

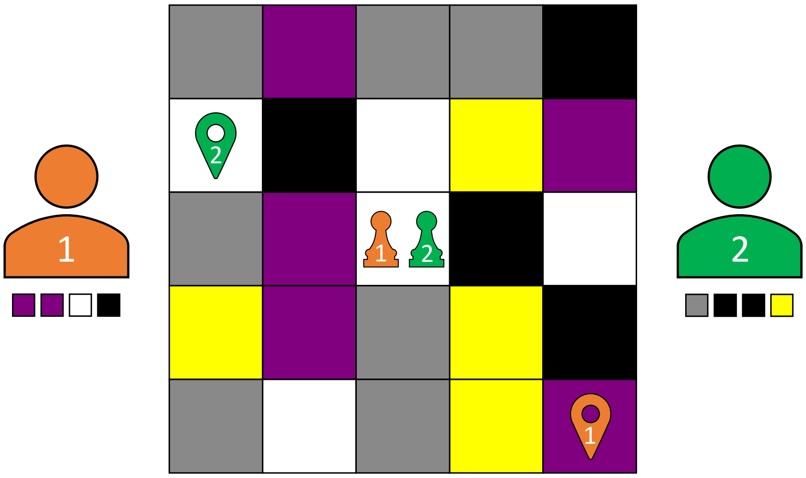


Figure 6.2: Illustration of the colored trails game used in the experiment. The chess pawns and numbered pins are players' game pieces and goal locations, respectively. The mini squares under players' illustrations denote their chip sets.

The game had two phases: a *negotiation phase* and a *movement phase*. In the negotiation phase, the two players could exchange their chips through negotiation, during which they alternately made *offers* to each other until an offer was accepted or a player withdrew from the negotiation. An offer describes a propositional redistribution of all the chips in the game. Maximally six offers could be made during the negotiation, i.e., a quota of three offers per participant. Once the six-offer limit was reached, the current turn player could no longer make a new offer but only accept the other player's offer or withdraw from the negotiation. If an offer were accepted, both players' chip sets would be redistributed as the offer indicates; otherwise, their chip sets would remain the initial distribution. Players with a ToM can infer their opponent's belief (i.e., goal location) from the negotiation process and, theoretically, have a better chance of reaching an agreement than players without a ToM [de Weerd et al., 2017]. Following the negotiation was the movement phase, during which players had unlimited time to move their game pieces until satisfied.

A player's score was calculated at the end of the game based on three criteria: (1) whether the game piece reached the goal location; (2) how many steps the game piece was moved toward the goal location; (3) how many chips were left. A player receives 100 points for each step made toward the goal location. Arriving at the goal location grants 500 points. Each unused chip adds 50 points.

6.2.2 Experiment Design

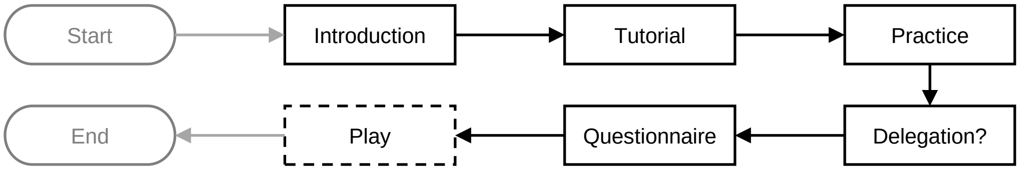
Initially, participants were informed that, during the experiment, they would play the colored trails game against a virtual agent and that they could win a very high bonus in addition to the standard monetary reward for their participation. The likelihood of winning the bonus was claimed to be positively correlated with their performance in the game. In fact, the bonus was fabricated to render the game outcomes consequential and induce the sense of criticality. Participants were remunerated with the standard reward and clarified the deception afterward.

The experiment procedure is illustrated in Figure 6.3. Before playing against the agent, participants were briefed on the game rules in an interactive tutorial and practiced the game by playing it against another virtual agent for five rounds. To facilitate the following discussion, I will use the term *trial agent* referring to the virtual agent for practicing, and the term *opponent agent* denoting the one to be contended with. During the practice, participants' view in the virtual environment was fixed and oriented toward the trial agent, as Figure 6.4 depicts. After the practice and before playing against the opponent agent, participants were offered an opportunity to delegate, i.e., letting the trial agent play against the opponent agent on their behalf. Once the delegatory decision was made, participants were asked to fill in a brief questionnaire[1] (cf. Table 6.1) consisting of two parts: (1) the *Trust in Automation Questionnaire* [Jian et al., 2000] for measuring the trial agent's trustworthiness and (2) two single items for measuring perceived task criticality and participants' intention to delegate. The questionnaire was answered on a seven-point Likert scale and additionally included two attention checks. The experiment was over once participants finished the questionnaire.

---

[1] The questionnaire was conducted directly in the virtual environment to minimize the influence of the break in presence [Putze et al., 2020].

The playing against the opponent agent was skipped, because it took place after the delegatory decision and the questionnaire, thus having no impact on the experiment results.



#	Step	Detail
1	Introduction	Participants are given an introduction of the experiment.
2	Tutorial	Participants are given a tutorial on the game.
3	Practice	Participants practice the game with the trial agent.
4	Delegation?	Participants decide whether to delegate.
5	Questionnaire	Participants answer a questionnaire.
6	Play	(This step was skipped) Participants or the trial agent plays against the opponent agent.

Figure 6.3: Flowchart illustrating the experiment procedure. The last step (i.e., "Play") is represented as a dashed rectangle because it was skipped.

Participants were divided into three groups, each practicing the game with a trial agent that bears a unique level of ToM. The different levels were implemented using the algorithm proposed in [de Weerd et al., 2017] based on orders of ToM (cf. Section 6.2.3).

6.2.3 Agent Model

The trial agent's model, implemented using the algorithm proposed by de Weerd et al. [2017], consisted of mainly three components: (1) the *belief* about whether an offer would be accepted by the opponent, (2) the *possibility* of a tile being the opponent's goal location, and (3) the *confidence* in a ToM order when the agent is capable of using multiple orders (e.g., a second-order ToM agent can also use the zero-order and first-order ToM). The belief was represented as a frequency table storing the acceptance rate of all the offers that could possibly appear in the game. The table's initial values are uniformly set to one, i.e., believing that the opponent would be happy to accept whatever offer was made. Such a belief

Table 6.1: Questionnaire items (attention checks excluded).

#	Item
1	Winning the bonus reward is a critical task to me.
2	I intend to delegate the succeeding games to the agent.
3	The agent is deceptive.
4	The agent behaves in an underhanded manner.
5	I am suspicious of the agent's intent, action, or outputs.
6	I am wary of the agent.
7	The agent's actions will have a harmful or injurious outcome.
8	I am confident in the agent.
9	The agent provides security.
10	The agent has integrity.
11	The agent is dependable.
12	The agent is reliable.
13	I can trust the agent.
14	I am familiar with the agent.

Measured factors: task criticality (#1), intention to delegate (#2), and trust (#3–14).

is obviously unreasonable, which makes it necessary to "educate" the agent via pre-training. In this experiment, the trial agent was initialized with a ready-to-use frequency table provided by de Weerd et al. [2017]. As the game proceeded, the agent would continuously update its belief based on the negotiation process. For example, when the agent's offer was not accepted, it would decrease the acceptance rate of this and other worse offers.

When negotiating with the opponent, the trial agent with a zero-order ToM (hereafter abbreviated as the ToM_0 agent) behaved solely based on its own belief. The ToM_1 agent additionally considered the opponent's belief and goal location by modeling the opponent as a ToM_0 agent. Since the opponent's goal location was hidden, the ToM_1 agent had to guess where the opponent's goal was. This ability to guess was achieved through the possibility component, where the agent gave each tile a percentage indicating the possibility of the tile being

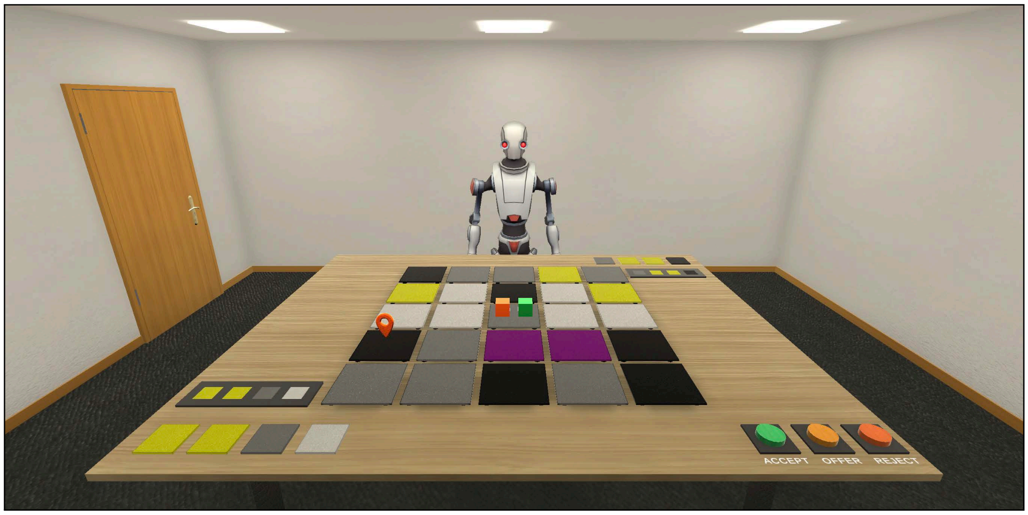


Figure 6.4: First-person perspective screenshot of the virtual environment used in the experiment. The robot is the trial agent. At the table center is the game board. The orange and green cubes on the center tile are participants' and the agent's game pieces, respectively. The orange floating pin points to participants' goal location (the black tile at the bottom left on the game board). At the table's bottom-right corner are three buttons, which constitute the player control. At the bottom-left corner, there are four bigger and four smaller squares. The bigger squares represent the chip set that participants currently have. The smaller ones on the gray platform denote participants' initial chip set. The agent's current and initial chip sets are displayed in the same layout at the table's top-right corner.

the opponent's goal. These percentage numbers were constantly updated by comparing the opponent's predicted and actual behaviors during the negotiation.

Although the ToM_1 agent was characterized by its first-order ToM, it could also use the zero-order ToM. During the negotiation, the ToM_1 agent would learn about the accuracy of the two ToM orders, which formed the confidence component for determining the best ToM order to use.

The ToM_2 agent functioned similarly to the ToM_1 agent, except that the opponent was modeled as a ToM_1 agent rather than a ToM_0 agent.

6.3 Results

Participants ($N = 150$) were recruited on Prolific and evenly assigned to the three groups, hereafter referred to as the ToM$_0$, ToM$_1$, and ToM$_2$ groups (cf. Section 6.2.2). Low-quality data were excluded from the results, such as those tendered by inattentive participants and speed runners. Ultimately, the data of 75 participants remained. Their demographics, delegation decisions, and game performance are given in Table 6.2.

Table 6.2: Participants' demographics, delegatory decisions, and performance.

Group	N	Age[1]	Gender	Delegation			Performance[3]	
				Yes	No	Rate[2]	Participant	Agent
ToM$_0$	23	M=33.4 SD=14.8	10 Females 13 Males	11	12	47.8%	531.7	629.1
ToM$_1$	25	M=29.7 SD=11.0	13 Females 12 Males	4	21	16.0%	571.2	544.8
ToM$_2$	27	M=29.3 SD=9.5	14 Females 13 Males	16	11	59.3%	469.3	598.1

1. M = mean, SD = standard deviation.
2. The rate refers to the proportion of participants who chose to delegate.
3. Performance was operationalized as the mean score of all participants or the mean score of the agent in the group.

The experiment results contradict the hypothesis that higher-order ToM agents are more likely to be delegated. Instead, there was a non-monotonic relationship between the delegation rate and ToM order. Only 16% of participants in the ToM$_1$ group chose to delegate, whereas the delegation rate of the other two groups was much higher at 50% or so. Participants' intention to delegate exhibited a similar pattern, where more participants in the ToM$_1$ group reported low-level intentions to delegate than the other two groups, but the difference was not statistically significant (Kruskal-Wallis test, $H = 4.844$, $p = 0.089$). Overall, the experiment results were unable to support the hypothesis and, when combined with other analyses discussed below, pointed to an alternative explanation that virtual agents' ToM only has a limited impact on users' choice and willingness to delegate to virtual agents.

Figure 6.5 illustrates participants' responses to the questionnaire. Most participants considered it critical to win the bonus ($H = 2.092$, $p = 0.351$), and their trusting attitudes were generally consistent ($H = 2.257$, $p = 0.323$).

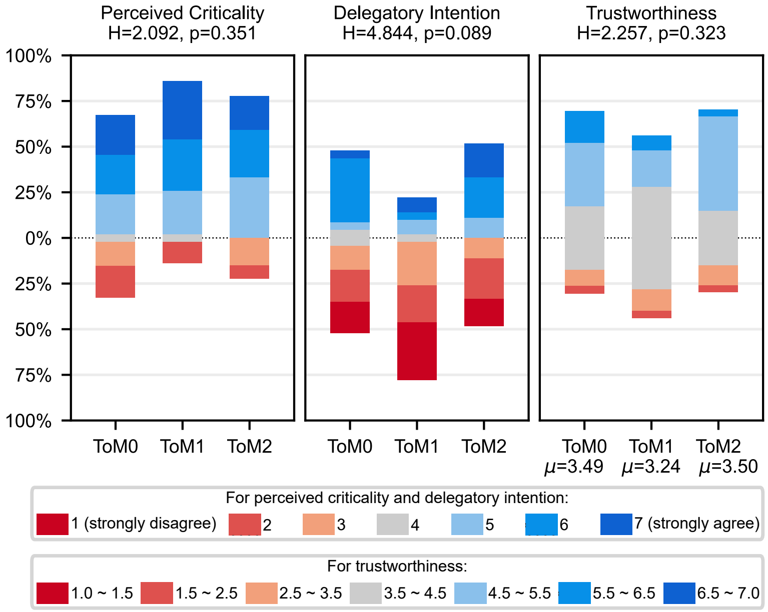


Figure 6.5: Participants' responses to the questionnaire. The colored bars indicate the proportions of different options, where 1 denotes "strongly disagree" and 7 denotes "strongly agree". The symbol μ in the tick labels refers to the mean value of the responses. The symbol H represents the test statistics of Kruskal-Wallis test .

An analysis of the entire data set (i.e., combining the three groups) revealed that performance was a relevant factor. As Table 6.3 shows, participants' intention to delegate was negatively correlated with their performance (operationalized as the mean score) in the game or the relative performance (operationalized as the agent's mean score minus participants' mean score). There was also a negative correlation between participants' performance and trust in the agent. However, neither trust nor the intention to delegate was correlated with the

Table 6.3: Correlation coefficients (Pearson's r).

	Trust	Delegation	Agent M.S.	Parti. M.S.	Score Diff.
Trust	1	-	-	-	-
Delegation	0.54^{*}	1	-	-	-
Agent M.S.	0.12	0.07	1	-	-
Parti. M.S.	-0.25^{*}	-0.29^{*}	0.50^{*}	1	-
Score Diff.	0.37^{*}	0.36^{*}	0.51^{*}	-0.49^{*}	1

* $p < 0.05$; M.S. = Mean Score; Parti. = Participant;
Score Diff. = Score Difference (agent's mean score minus participants' mean score)

agent's performance. Additionally, a strong-effect positive correlation between agent trustworthiness and the intention to delegate was observed, corroborating with other studies [Aggarwal and Mazumdar, 2008; Leana, 1987; Lubars and Tan, 2019; Stout et al., 2014].

6.4 Discussion

The correlation analysis above indicated that participants were egocentric and made delegation decisions mainly based on their own performance or relative performance. Such a self-centric stance might have rendered ToM a low-priority and consequently less significant factor. Indeed, losses and gains were an intuitive and reliable source of information for making delegation decisions. Comparatively, the perception of the trial agent's ToM was less straightforward, not to mention that such perception could be overlooked due to algorithmic opacity. Moreover, the focus on performance suggested that the relationship between participants and the trial agent was similar to that between supervisors and supervisees (e.g., managers and subordinates). ToM may only have a limited impact in a supervising relationship than peer-like or collaborative one [Hanna and Richards, 2018].

While participants' different performance may account for their delegation decisions, where the difference originates from is another question. Analysis showed that the frequency of successful negotiation (i.e., reaching an agreement) was positively correlated with both participants' score ($r = 0.645, p < 0.001$) and the agent's score ($r = 0.584, p < 0.001$). More specifically, participants' score depended on how many of their offers were accepted by the agent ($r = 0.593, p < 0.001$) and, likewise, the agent's score was contingent on the number of its offers accepted by participants ($r = 0.408, p < 0.001$). However, for participants, their behavior (i.e., accepting the agent's offers or withdrawing from the negotiation) had a limited impact on their score ($r = -0.072, p = 0.538$), and the same holds for the agent ($r = 0.082, p = 0.483$). In the experiment, with the agent's ToM order increasing from zero to two, the mean number of offers the agent accepted changed non-linearly from 1.22, over 2.64, to 1.56. This non-linearity may explain participants' different performance across the three groups and, in turn, was manifested in the non-monotonic relationship between ToM orders and delegation rates (cf. Table 6.2). The mutual dependency may result from players' rationality, as an offer is often more beneficial to the proposer than the responder.

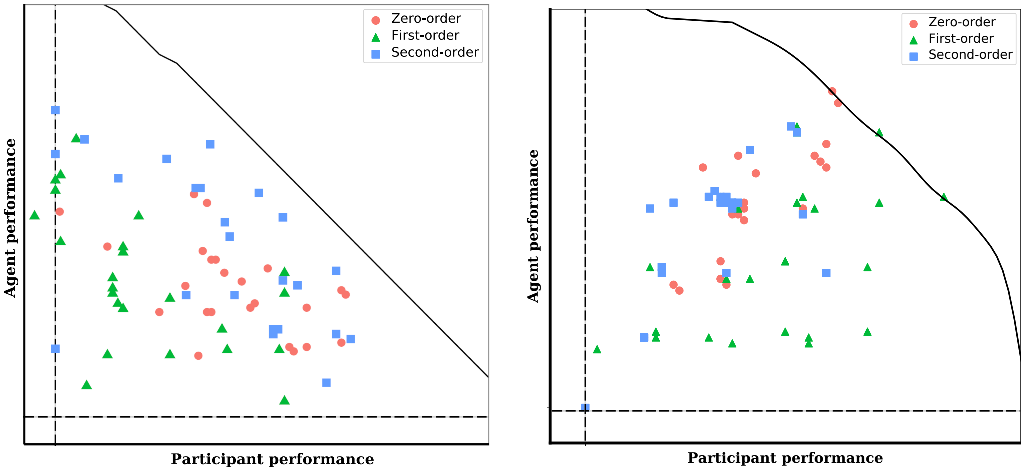
This experiment employed a different game configuration (game board color distributions, initial chip sets, and goal locations) from the original study [de Weerd et al., 2017] because the original configuration was not disclosed. Some other changes (detailed in Table 6.4) were made to facilitate investigating human-agent interaction over testing the algorithm performance. Consequently, the two settings produced different results. As Figure 6.6 illustrates, players' performance was positively correlated in this experiment but negatively correlated in the original study. The divergence was likely caused by different intensities of *conflict of interest*, i.e., the number of chips that both players demanded. In the original setting, an intense conflict of interest might exist, and the two players had to compete over limited resources. In this experiment, however, the conflict of interest was relatively moderate, in which case a successful negotiation was more likely to be mutually beneficial than a compromised solution to the dilemma.

Implication for the Conceptual Model (cf. Chapter 3)

ToM is a crucial skill for interpersonal communication and relationships. Its limited impact on delegation decreases the significance of the affective dimension in users' delegation decisions.

Table 6.4: Experiment setting comparison.

Setting	This experiment	The original study
Participant count	75	27
Experiment type	Between subjects	Within subjects
Game rounds	5	8
Initiator	Participant	Alternating
Time limit	No	Yes
Belief reset	No	Yes



(a) The original study.

(b) This experiment.

Figure 6.6: Player performance as increased scores. Figure 6.6a is a screenshot taken directly from [de Weerd et al., 2017]. Figure 6.6b presents the results of this experiment, adapting the layout of Figure 6.6a to facilitate comparison. The dashed lines denote the maximal score that a player can obtain only with the initial chip set. The solid black line shows the Pareto efficient outcomes.

Limitations

A major limitation is the lack of control for performance. As indicated by the experiment results, the impact of performance on participants' delegation decisions was prominent, insofar as it might have overshadowed the impact of ToM. To alleviate this problem, one may consider manipulating the game outcomes to ensure that performance-related variables, such as scores or win/loss, are generally at the same level for all participants.

The communication between participants and the agent was severely constrained in the experiment, during which the only message exchanged between them was the offers. This constitutes another major limitation since virtual agents have and feature the ability to utilize human communication, such as verbal or body language, which can be used to better present their negotiation and ToM skills. Allowing virtual agents to communicate with users in these channels may make the impact of ToM more pronounced to be observed.

Although there is a huge space for players to utilize their ToM in the colored trails game, the game is too complicated for participants to grasp it rapidly, particularly for a remote experiment. Participants might have already lost patience during the pre-experiment tutorial or felt a heavy cognitive workload when playing the game. Consequently, participants might speedrun the experiment and tender relatively low-quality data. For future studies, one may consider using other simpler games like the Price Game used in [Mou et al., 2020]. Besides, ToM is potentially more influential in collaborative scenarios than in relatively competitive games as in the colored trails game. Therefore, one may also consider employing a fully collaborative context to increase the possibility of observing a significant effect.

Due to the complex nature of ToM, findings derived from this experiment can only be generalized to other contexts to a limited extent. For example, the ability to infer emotion and intention can be equally important as belief inference for a virtual agent's ToM, but these two aspects were not considered in the experiment. Furthermore, there are other ways to theoretically categorize ToM apart from ToM orders (e.g., an often-used criterion is Sally-Anne test). The findings may not apply to those contexts based on a different ToM architecture. Thus, more studies focusing on other aspects or architectures of ToM are needed to fully understand its impact on delegation.

6.5 Chapter Summary

This chapter presented an empirical study investigating whether and how a virtual agent's ToM influences delegation to the agent. A between-subjects experiment was conducted to compare three virtual agents with different levels of ToM. The results showed that the agents' ToM might only have a limited impact on participants' delegatory decisions and intention. Nevertheless, performance was a relevant factor in the intention to delegate, and participants' performance was particularly influential possibly due to their ego-centrism. Thus, when designing virtual agents for carrying out critical tasks, developers may consider refraining from using ToM-resemblance features and focusing on balancing user performance perception.

The next chapter looks into the notion of technological immersion and explores its impact on delegation decisions based on the results of two experiments.

Chapter 7

Delegation to Virtual Agents in Immersive Settings

Portions of the content presented in this chapter were published in [Sun and Botev, 2023].

Virtual agents are inherently simulated, so interaction with virtual agents must be mediated by a technical device. Commonly employed devices include desktop computers, laptops, tablets, and smartphones [Abdullah et al., 2018; Gan et al., 2022]. These devices mainly feature a two-dimensional panel display to (mostly visually) mediate and represent an environment as if users were viewing it through a window. However, this looking-through-a-window way of interaction is not ideal for delivering immersive interactive experiences, despite being widely used and offering various advantages [Chowdhury et al., 2021; Klingenberg et al., 2020]. To overcome this limitation, researchers and companies have devoted a substantial amount of effort over the past decades to developing immersive media devices, particularly head-mounted displays for VR and AR. With more advanced features such as stereoscopic displays and panoramic field of regard, this class of devices is capable of immersing users in a "technology-driven environment with the possibility to actively partake and participate in the information and experiences dispensed by the generated world" [Perkis et al., 2020].

A key difference between the two classes of media devices mentioned above lies in their levels of *technological immersion*, a term that describes a device's capacity to mediate and represent an environment in a way that matches human perception of the physical world. Unlike the subjective feeling of being immersed, technological immersion is “a quantifiable description of a technology” [Slater and Wilbur, 1997]. Following these definitions, reading a compelling comic book magazine is psychologically more immersive than watching a boring movie on an 85-inch screen. Yet, the screen is technologically more immersive than the magazine because of its wide field of regard, surround sound, and more. The prevailing standard for measuring technological immersion is the five dimensions proposed by Slater and Wilbur [1997], i.e., inclusiveness, extensiveness, surroundingness, vividness, and matching:

> ***Inclusive*** (I) indicates the extent to which physical reality is shut out. ***Extensive*** (E) indicates the range of sensory modalities accommodated. ***Surrounding*** (S) indicates the extent to which this virtual reality is panoramic rather than limited to a narrow field. ***Vivid*** (V) indicates the resolution, fidelity, and variety of energy simulated within a particular modality (for example, the visual and color resolution) [...] ***Matching*** requires that there is match between the participant’s proprioceptive feedback about body movements, and the information generated on the displays. [Slater and Wilbur, 1997]

Various effects of high-level technological immersion have been documented in the literature. For instance, there is ample evidence that people experience a stronger sense of presence when using head-mounted displays rather than desktops [Buttussi and Chittaro, 2018; Cummings and Bailenson, 2016; Guimarães et al., 2020; Kim et al., 2014; Makransky et al., 2019; Moreno and Mayer, 2004; Ochs et al., 2019; Reinhard et al., 2022]. Immersive media devices can also better support 3D-intensive tasks, such as spatial learning [Krokos et al., 2019; Pollard et al., 2020] and object detection [Henderson and Feiner, 2009; Odenthal et al., 2014; Reinhard et al., 2022], when compared with less immersive devices, as in desktops or smartphones. Psychological studies often report a positive correlation between technological immersion and stronger emotional responses [Kim et al., 2014; Pallavicini and Pepe, 2019; Simon and Greitemeyer, 2019] or social effects [Chowdhury et al., 2021; Ochs et al., 2019; Pejsa et al., 2017]. In education, students can benefit from being taught via technologically immersive devices with better knowledge retention and transfer [Klingenberg et al., 2020]. An overview of the effects of technological immersion is provided below.

Positive Effects:

- Stronger sense of presence and co-presence [Ochs et al., 2019].
- Higher-level psychological immersion and more positive emotions in video game playing [Pallavicini and Pepe, 2019].
- Higher-level emotional arousal in the Stroop task [Kim et al., 2014].
- Higher-level physiological and sexual arousal when viewing pornographic video materials [Simon and Greitemeyer, 2019].
- Better improvement of social skills in persons with the *autism spectrum disorder* after *virtual social skill interventions* [Miller and Bugnariu, 2016].
- Lower-level implicit bias toward persons with disabilities after experiencing a *disability simulation* [Chowdhury et al., 2021].
- Better knowledge retention and transfer in students educated using *generative learning strategies* [Klingenberg et al., 2020].
- Higher recall accuracy in using a virtual *memory palace* [Krokos et al., 2019].
- Higher accuracy in detecting assembly errors in a model [Odenthal et al., 2014].
- Better performance in object recognition and discrimination after spatial learning in a virtual environment [Pollard et al., 2020; Reinhard et al., 2022].

Negative Effects:

- Higher probability of the Uncanny Valley effect [Hepperle et al., 2020], fatigue [Madeira et al., 2022; Plechatá et al., 2019], and the *simulator sickness* [Cao et al., 2020].
- Higher cognitive load and worse learning outcome after learning science through virtual learning simulations [Makransky et al., 2019].

Neutral Effects:

- Similar performance in 3D navigation [Feng et al., 2022; Sousa Santos et al., 2008] and 3D object manipulation [Madeira et al., 2022].
- Similar performance in the spatial understanding of a complex 3D virtual model, such as livers or pyramids [Hattab et al., 2021].
- Similar performance in cognitive recovery in stroke patients after a virtual exercise [Gamito et al., 2014].
- Similar performance in knowledge acquisition after learning digitally [Buttussi and Chittaro, 2018; Olmos-Raya et al., 2018].
- Similar levels of engagement in video game playing [Cao et al., 2020].
- Similar levels of intention to reuse a virtual shopping environment after experiencing it [Peukert et al., 2019].

The ample evidence presented above points to the possibility that technological immersion may also have an impact on delegation decisions to virtual agents. To test whether the impact exists, two experiments were conducted – a within-subjects and a between-subjects experiment– to observe and compare users' delegation to virtual agents in different conditions and levels of technological immersion. Before detailing the experiments, this chapter further clarifies the concept of immersion in Section 7.1.

7.1 Immersion Definitions

In research related to virtual environments, the term "immersion" is commonly used to denote the situation of users being surrounded or enclosed by an experience, an interaction, or a mediated reality, as the quote below describes.

> Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by

> a completely other reality, as different as water is from air, that takes over all of our attention, out whole perceptual apparatus. [Murray, 2017]

Although the general idea behind the term is clear, its practical definition has been diverging in academia. In their recent review article, Nilsson et al. [2016] identified numerous definitions, including *system immersion* (i.e., technological immersion), *sensory immersion, perceptual immersion, fictional immersion, psychological immersion, narrative immersion, imaginative immersion, systemic immersion, strategic immersion, tactical immersion, ludic immersion*, and *challenge-based immersion*. Among them, technological immersion constitutes a unique perspective that views immersion objectively as a description of technologies, whereas other definitions predominantly take a psychological approach and regard immersion as a subjective construct, such as feeling, attention, or mental experience. The prevailing standard for measuring technological immersion is the five dimensions proposed by Slater and Wilbur [1997], namely inclusiveness, extensiveness, surroundingness, vividness, and matching. The more capable a media device is in each dimension, the more immersive the device is, and the more the device "delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities" [Slater, 2003]. Apart from technological immersion, Nilsson et al. [2016] classified the other definitions into three categories.

1. Sensory and perceptual immersion are closely related to technological immersion because both of them regard immersion as the subjective experience of an individual's senses being enclosed by a technology-driven environment. According to these two definitions, immersion increases when, for example, "large screens close to the players' eyes and powerful sounds easily overpower the sensory information coming from the real world" [Ermi and Mäyrä, 2005] or when the physical world is hidden from users "by the use of goggles, headphones, gloves, and so on" [McMahan, 2013]. The major difference between technological and sensory/perceptual immersion is that technological immersion objectively describes the technical configuration of a hardware device,

whereas sensory/perceptual immersion refers to the extent to which users feel like their senses are enveloped by the hardware device.

2. The technology-driven definitions mentioned above cannot explain why reading a compelling comic magazine feels more immersive than watching a boring documentary, despite books being a relatively primitive form of media. This question is answerable if one follows another category of immersion definitions, including psychological, narrative, fictional, and imaginative immersion. These definitions are highly similar and all emphasize the extent to which individuals are involved or absorbed in a narrative. Technology degrades to an insignificant factor, whereas the storytelling elements such as plots and characters become paramount for immersion.
3. Immersion may also occur without narratives and advanced technologies, such as when playing Tetris on one of the early mobile phones. This type of immersion has been named systemic, strategic/tactical, and challenge-based immersion, but is more widely known under the term “engagement”. These definitions argue that immersion originates from individuals trying to utilize their mental or sensorimotor skills intently to achieve a certain performance in a task. During the process, individuals are highly attentive to the task at hand and thus become absorbed in the task.

7.2 Experiment 1 – Within-Subjects

Following a within-subjects design, Experiment 1 ($N = 30$) compared delegation to virtual agents across three different levels of technological immersion. The experiment was on-site and approved by the ethics review panel at the University of Luxembourg under file number ERP 22-049 DELICIOS.

7.2.1 Methodology

During the experiment, participants were tasked to play the Colonel Blotto game. The original Colonel Blotto game involves a martial scenario between two warring countries. In this experiment, the game was recontextualized with a commercial scenario, where two beverage companies compete for three cities' beverage markets. Participants acted on behalf of one company and played against a virtual agent that controlled the other company. Each company initially had ten-liter concentrate of its beverage product and could distribute them to the three cities. The minimal division unit was one liter. The concentrate allocated for each city would be further diluted, bottled, and sold in the city. A company wins a city's market if the company distributes more concentrate to the city than the other company. A player wins the game if the player's company wins two markets. Players had unlimited time to set their distributions, which were not shown until all the players confirmed theirs.

With the ten-liter cap, the number of different distributions that a player could possibly make was limited within a suitable range that was neither too big[1] nor too small[2]. Findings derived from such a moderate-complexity setting are arguably more generalizable than low- or high-complexity settings since, in real-world cases, available strategies are often not binary but constrained to a finite set.

Procedure

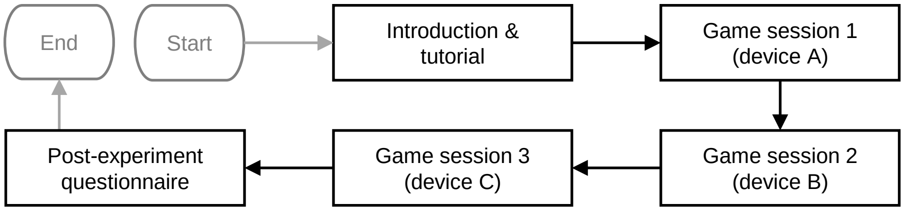
As illustrated in Figure 7.1, after receiving an introduction and tutorial on the game, participants were asked to finish three consecutive game sessions and then complete a post-experiment questionnaire.. Each game session, consisting of six consecutive game rounds, followed the same procedure (cf. Figure 7.2) but was played at a unique level of technological immersion.

To render the game critical, participants were led to believe that their monetary reward depended entirely on their game performance. The first three rounds of a game session constituted a training phase, where participants practiced the game by playing it against a *trial agent* (cf. Figure 7.3a). The outcomes of these three rounds had no impact on their reward. In the remaining three rounds, participants played against an *opponent agent* (cf. Figure 7.3b) whose strategy was claimed to be different from the trial agent but, in fact, the same. The outcomes of the remaining three rounds pertained to participants' rewards: for each round they

---

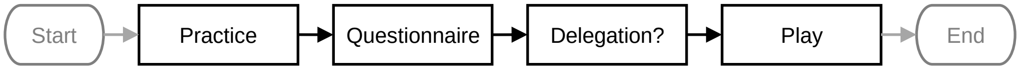
[1] For example, 100 liters of concentrate would allow for thousands of possible distributions.

[2] For example, the Prisoner's Dilemma provides players with only two available options.



#	Step	Detail
1	Introduction & tutorial	Participants are given an introduction and then a tutorial on the game.
2	Game session 1 (device A)	Participants play a game session using device A.
3	Game session 2 (device B)	Participants play a game session using device B.
4	Game session 3 (device C)	Participants play a game session using device C.
5	Post-experiment questionnaire	Participants answer a post-experiment questionnaire.

Figure 7.1: Flowchart illustrating the procedure of Experiment 1.



#	Step	Detail
1	Practice	Participants practice the game with the trial agent in the first three rounds.
2	Questionnaire	Participants answer the condition-wise questionnaire.
3	Delegation?	Participants decide whether to delegate.
4	Play	Participants or the trial agent plays against the opponent agent.

Figure 7.2: Flowchart illustrating the game session.

won or lost, their reward would be increased or decreased by a small amount of money, respectively. The reward could not be deducted to negative and was minimally zero. The reward would not change if a game round ended with a draw. The deception was disclosed and explained by the end of the experiment. All the participants received the same standard compensation.

In between the first and remaining three rounds, participants filled in a brief questionnaire (cf. Table 7.1) within the mediated environment. After finishing the questionnaire, participants were offered an opportunity to delegate the playing of the remaining three rounds to the trial agent. If a participant chose to delegate, the trial agent would take over complete control and autonomously play the remaining rounds on the participant's behalf.

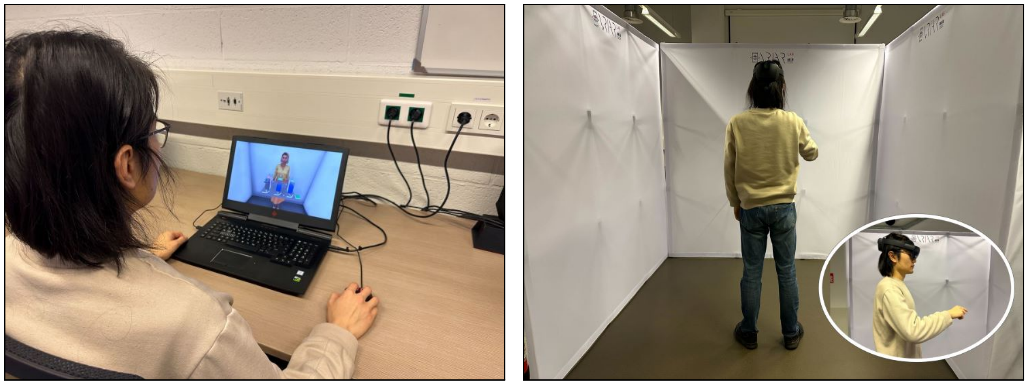

(a) The trial agent. **(b)** The opponent agent.

Figure 7.3: Virtual agents used in Experiment 1.

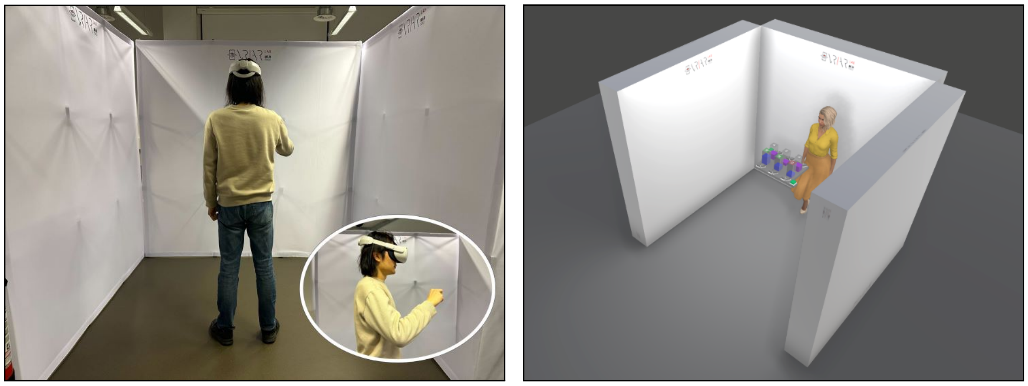
Following a within-subjects experimental design, the three game sessions were played using different media devices that vary in their technological immersion levels (cf. Figure 7.4 and the next paragraphs). The within-subjects design helps to control for the impact of individual differences in, for example, demographics or experience with similar technologies. To mitigate the influence of potential carryover effects, the order of the three devices was randomized for each participant using a Latin square design. To further minimize its influence, the virtual agents in each game session –including both the trial and opponent agents– were claimed to be unique and have different behaviors and strategies from the agents in other game sessions, though all of them were actually driven by the same rules. The experiment was over once participants finished the last session, after which a post-experiment questionnaire (cf. Table 7.2) was administered.

Variables, Conditions, and Measures

The independent variable was the media device's technological immersion level. Based on the five dimensions mentioned previously, three distinctive types of devices were selected as the three conditions of the independent variable, including



(a) Low-immersion setup. **(b)** Medium-immersion setup.



(c) High-immersion setup. **(d)** The virtual environment.

Figure 7.4: The physical setting of each condition in Experiment 1. The devices used in the low-, medium-, and high-immersion conditions were a laptop, a Microsoft HoloLens 2, and a Meta Quest 2, respectively. Participants were placed in a confined space amid three large neutral walls. The virtual environment (cf. Figure 7.4d) used for the low- and high-immersion setups replicates the physical space on a 1:1 scale.

an OMEN gaming laptop, a Microsoft HoloLens 2, and a Meta Quest 2.

Laptop (low-immersion condition). Environments mediated by the laptop are only marginally inclusive because, in typical cases, its 17-inch screen can only occlude a fraction of the physical reality. The laptop's extensiveness is high since it utilizes both visual and auditory channels to communicate with users. The screen can display 81 pixels per inch at the Full-HD level, offering a relatively high level of vividness but falling short of surroundingness due to its limited field of regard. Compared with the other two devices, the laptop is also inferior in matching as it only tracks some simple user inputs, such as mouse movements and keyboard strokes.

HoloLens (medium-immersion condition). The screen of the HoloLens is transparent, due to which environments mediated by it are minimally inclusive. The HoloLens has a similar level of extensiveness as the laptop but displays less vivid images rendered at 42 pixels per degree. Nevertheless, the HoloLens has a much higher level of surroundingness than the laptop due to its panoramic field of regard and stereoscopic display. The HoloLens also outperforms the laptop in matching with its higher-resolution head and hand tracking.

Quest (high-immersion condition). Environments mediated by the Quest are highly inclusive as the Quest has a dedicated design for hiding the perception of real-world environments from users. The Quest is comparable to the HoloLens in terms of extensiveness and matching, but it offers a higher level of surroundingness with its much larger field of view at 97 degrees compared to the 37 degrees provided by the HoloLens. As a side effect of its large field of view, images displayed by the Quest are rendered less vividly at only 19 pixels per degree.

The dependent variable was participants' delegation decisions in the three game sessions.

In addition to the independent and dependent variables, some other factors were also measured using questionnaires to learn about participants' decision-making process and to check for potential alternative explanations. These factors were measured using questionnaires on a 101-point scale that ranged from 0 to 1, with 0 and 1 denoting "strongly disagree" and "strongly agree", respectively. The condition-wise questionnaire (cf. Table 7.1) inquired about three factors related to delegation, including the trustworthiness and competence of the trial agent and the perceived workload during the game. Notably, the condition-wise questionnaire also probed participants' subjective feeling of immersion, allowing us to examine its relationship with technological immersion. Since participants needed to repeat answering the condition-wise questionnaire three times in a short period of approximately 5–10 min, each of the four factors mentioned above was measured using a single item for simplicity and avoiding boredom and frustration in participants, though at the cost of decreased reliability and validity. Nevertheless, the four factors are relatively straightforward as opposed to more complex constructs such as presence, owing to which the single-item measure is arguably sufficient for comparing these factors across the three conditions [Allen et al., 2022; Ang and Eisend, 2018]. Participants were additionally asked to choose the most trustworthy trial agent among the three conditions. The post-experiment questionnaire (cf. Table 7.2) checked on some other delegation-related factors and asked participants to directly compare the trustworthiness of the trial agent in different game sessions.

Table 7.1: The condition-wise questionnaire.

#	Item
1	The agent in the last rounds was trustworthy.
2	The agent in the last rounds was competent in the game.
3	I felt cognitively overloaded in the last rounds.
4	I felt immersed or involved in the game over the last rounds.

Measured factors: trustworthiness (#1), competence (#2), workload (#3), and psychological immersion (#4). Regarding item 4, research suggests that people can “reliably reflect their own immersion in a single question” [Jennett et al., 2008].

Table 7.2: The post-experiment questionnaire (Experiment 1).

#	Item
1	The game outcome is important to me.
2	The agents felt controllable.
3	I must account for the game outcome.
4	Which agent did you find most trustworthy?

Measured factors: criticality (#1), controllability (#2), accountability (#3), and trustworthiness (#4).

Agent Design

An individual's decision on delegation can be significantly influenced by performance-related information [Leana, 1987; Sun et al., 2022; Yukl and Fu, 1999]. Thus, to control for the impact of performance, an agent (either the trial agent or the opponent agent) was allowed to cheat when the agent played the role of the participants' rival in the game. When cheating, the agent was secretly informed of the concentrate distribution on the participants' side. With the information, the rival could win or lose certain rounds deliberately. For every three rounds in a game session, the rival randomly chose one of the following pre-defined sequences to enact: (1) win, lose, lose (the rival deliberately wins the first round and then loses the second and third rounds); (2) lose, win, lose; (3) lose, lose, win; (4) win, draw, draw; (5) draw, win, draw; or (6) draw, draw, win. For example, within the first three rounds of a game session, the trial agent may choose to win the first round and lose the second and third rounds, whereas, within the remaining three rounds, the opponent agent may deliberately lose the first and second rounds and win the third round. The different sequences all result in the same payoff: participants receive the money equivalent to winning one round.

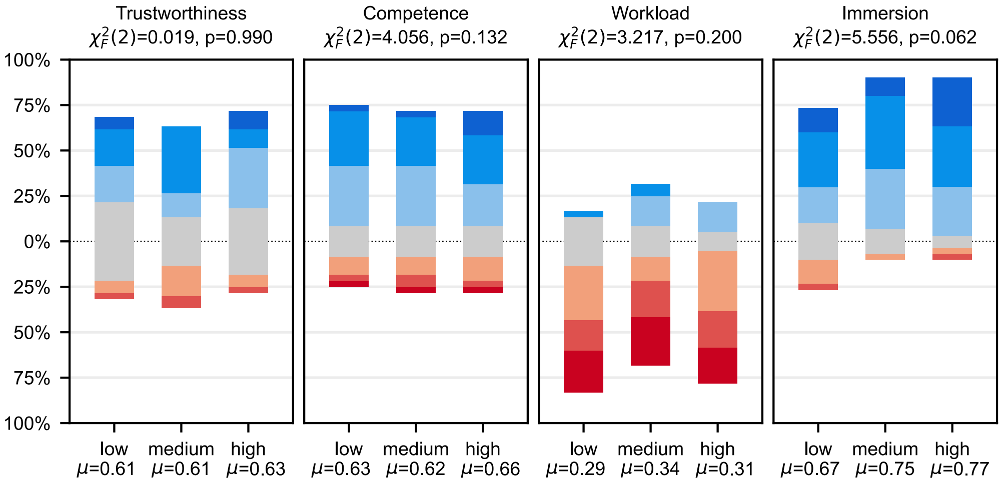
The virtual agents were embodied as virtual humans in light of our previous finding that users prefer to interact with humanlike virtual agents in critical scenarios involving delegation (cf. Chapter 5). The virtual human characters used in the experiment were generated using Character Creator 4, a software application that provides an integrated solution for creating high-quality digital characters. The characters were lightly animated to make the interaction with them feel natural for participants. The animation combined a full-body idle animation clip with randomly generated simple facial motions, such as eye blinks or subtle movements of facial muscles. The trial and opponent agents were invariably embodied in a virtual woman and man across the three conditions, respectively. Although an agent's gender can bias people's delegation to the agent [Akinola et al., 2018; Leana, 1987; Payne et al., 2013], its effect is arguably not potent enough to overshadow the effect of technological immersion. The appearances of the virtual agents in different game sessions were identical, except that the color of their shirts was varied to visually emphasize that they were not the same agent. For example, the trial and opponent agents in one game session may both wear yellow shirts (cf. Figure 7.3), whereas, in another game session, they may both wear blue shirts.

7.2.2 Results

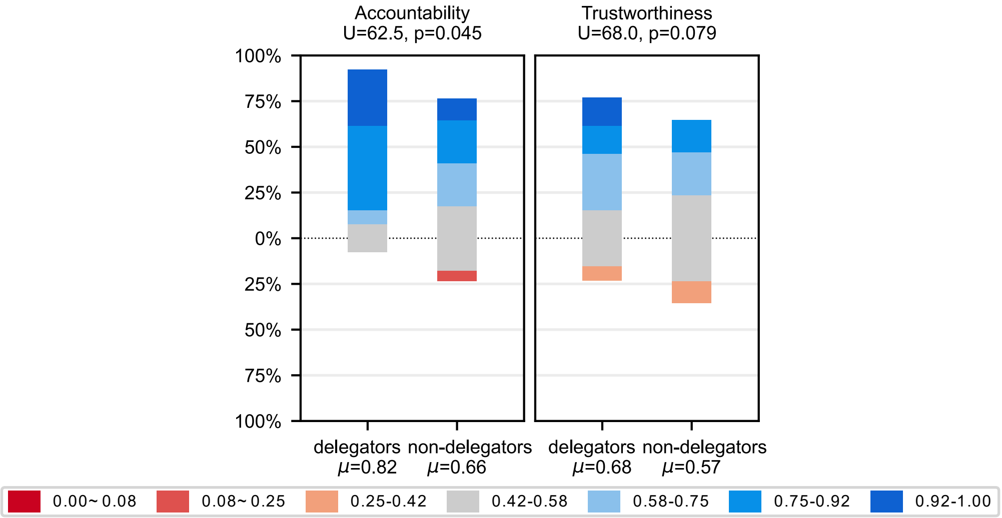
The participants (N = 30, 15 males and 15 females, mean age = 26.4) were recruited by disseminating a call to the student body at the University of Luxembourg. The exclusion criteria were that participants must be healthy adults without any motor handicaps and previous symptoms of seizures. The experiment results showed that participants were generally reluctant to delegate to the trial agent. Only four participants chose to delegate in the low-immersion condition, whereas this number increased to seven in the medium- and high-immersion conditions. There was no statistically significant difference in participants' delegation decisions among the three conditions (Cochran's Q test, $Q(2) = 2.571, p = 0.276$). These results cannot support the hypothesis that technological immersion influences delegation decisions to virtual agents.

Figure 7.5a summarizes participants' responses to the condition-wise questionnaire. The assessments of the trial agent's trustworthiness were almost identical under different conditions (Friedman test, $\chi^2_F(2) = 0.019, p = 0.990$). A similar pattern can be found in the assessments of the trial agent's competence, except for a slightly higher variance ($\chi^2_F(2) = 4.056, p = 0.132$). The perceived workload was generally low during the game sessions ($\chi^2_F(2) = 3.217, p = 0.200$), whereas the subjective feeling of immersion was generally high. Participants felt less immersed in the low-immersion condition than in the other two conditions ($\chi^2_F(2) = 5.556, p = 0.062$), while the medium- and high-immersion conditions were psychologically immersive to similar extents.

The responses to the post-experiment questionnaire showed that participants generally considered the game outcomes critical; the mean value of their assessments of the task criticality was 0.705 (SD = 0.209; Shapiro-Wilk test, $W = 0.903, p = 0.010$). Corresponding to the high-level criticality, the perceived accountability for the game outcomes was also high, with a mean value of 0.731 (SD = 0.208; $W = 0.905, p = 0.011$). The controllability assessments were



(a) Participants' responses to the condition-wise questionnaire.



(b) Delegators' and non-delegators' perceived accountability and trustworthiness.

Figure 7.5: Participants' questionnaire responses categorized into seven bins, as illustrated in the legend. The colored bars indicate the proportions of different bins, where 0 denotes "strongly disagree", and 1 denotes "strongly agree". The symbol μ in the tick labels refers to the mean value of the responses. The symbols χ^2_F and U represent the test statistics of Friedman test and Mann-Whitney U test, respectively.

positive (i.e., the trial agent was controllable) but had a relatively low mean value of 0.630 (SD = 0.226; $W = 0.929, p = 0.047$). When participants were asked to choose the most trustworthy trial agent among the different conditions, eight of them picked the agent in the low-immersion condition, whereas this number in the medium- and high-immersion conditions grew to 12 and 10, respectively. This result conformed to participants' tendencies to delegate, where more participants chose to delegate in the medium- and high-immersion conditions than in the low-immersion condition.

Of all the participants, 17 were *non-delegators*, i.e., those who did not delegate in any of the game sessions. The other 13 participants were *delegators* who chose to delegate in at least one of the game sessions: ten participants delegated in one session, one participant delegated in two sessions, and two participants delegated in all three sessions. Delegators and non-delegators exhibited some notable differences in their questionnaire responses (cf. Figure 7.5b). Delegators reported a considerably higher level of perceived accountability for the game outcomes than non-delegators (Mann-Whitney U test, $U = 62.5, p = 0.022$). This result corroborates an earlier finding that there is a positive correlation between users' willingness to delegate a task to software agents and users' perceived accountability for the task outcomes [Stout et al., 2014]. Furthermore, when averaging each participant's assessments of trustworthiness in the three conditions, delegators were found to put a higher level of trust in the trial agent than non-delegators ($U = 68.0, p = 0.039$), which is also consistent with other studies showing a positive correlation between trust and delegation. Most delegators (9 out of 13) picked the trial agent in the medium-immersion condition as the most trustworthy one, whereas only one and three delegators picked the trial agent in the low- and high-immersion conditions, respectively. The preference for the medium-level immersion pointed to the possibility of using mixed-reality environments to increase users' trust in virtual agents.

A correlation analysis of the data revealed that the perceived trustworthiness and competence of the trial agent were positively correlated, except that the correlation in the medium-immersion condition was close to statistical significance (for the low-immersion condition, $r = 0.374, p = 0.042$; for the medium-immersion condition, $r = 0.352, p = 0.055$; for the high-immersion condition, $r = 0.466, p = 0.010$). The analysis also identified some other inconsistent correlations across the three conditions. For example, there was a positive correlation between perceived competence and psychological immersion in the high-immersion condition ($r = 0.388, p = 0.034$), yet this correlation was absent in the low-immersion condition ($r = 0.167, p = 0.376$) and medium-immersion condition ($r = 0.023, p = 0.902$). The correlation between perceived competence and workload was statistically significant in the medium-immersion condition ($r = 0.435, p = 0.016$) but insignificant in the low-immersion condition ($r = 0.117, p = 0.538$) and high-immersion condition ($r = 0.335, p = 0.071$). When data from the three conditions were aggregated, a positive correlation was found between perceived trustworthiness and competence ($r = 0.399, p < 0.001$), consistent with other findings on the relevance between trust and competence-related factors such as ability [Mayer et al., 1995] or performance [Lee and Moray, 1992]. Perceived workload and competence were also positively correlated ($r = 0.305, p = 0.003$), which suggested that participants considered the trial agent more competent when they felt like having a heavy workload.

7.2.3 Limitations

The operationalization of the dependent variable comprised only a binary choice on delegation. This coarse-grained measure may hardly capture the effect of the independent variable when the effect is small. Thus, one may consider employing more fine-grained measures to detect nuanced changes in participants' delegatory behavior or intention, such as letting participants make a series of delegation decisions or, alternatively, using questionnaires to assess participants' intention to delegate as in the case of [Stout et al., 2014].

Similar to the issue discussed above, the questionnaires used in the experiment might only have a low level of reliability and validity since they were overly simplified to cram the repeated measures in a short period of time. A few more questions could be added to the questionnaires to improve their validity and reliability without significantly increasing the risk of boring or frustrating participants. For example, items from the Trust in Automation Questionnaire [Jian et al., 2000] could be adapted into the condition-wise questionnaire.

There was a flaw in the dynamic rewarding mechanism that participants were led to believe at the beginning of the experiment. As a convention and regulation of user studies at the University of Luxembourg, participants are usually rewarded with a certain amount of money. In this experiment, the setting that the monetary reward could be deducted to zero might have been a giveaway. To make the deception more believable, one may consider claiming that there is a basic reward in addition to the dynamic one.

The within-subjects experiment design constitutes a potential limitation because participants' delegatory decisions could be influenced by their experience of previous game sessions. Although specific measures such as the Latin square design were taken to minimize the carryover effect, its influence arguably cannot be eradicated, especially when examining high-level cognitive processes such as trust or decisions on delegation. Studies employing different designs (e.g., between-subjects experiments) are needed to fully explore the relationship between technological immersion and delegation to virtual agents.

7.3 Experiment 2 – Between-Subjects

Considering the limitations of Experiment 1 and aiming to test the hypothesis again using a different approach, a follow-up experiment was conducted (Experiment 2, $N = 30$), which following a between-subjects approach to compare delegation to virtual agents in two conditions of distinctive technological immersion levels. The experiment was on-site and approved by the ethics review panel at the University of Luxembourg under file number ERP 23-037 DELICOS3.

7.3.1 Methodology

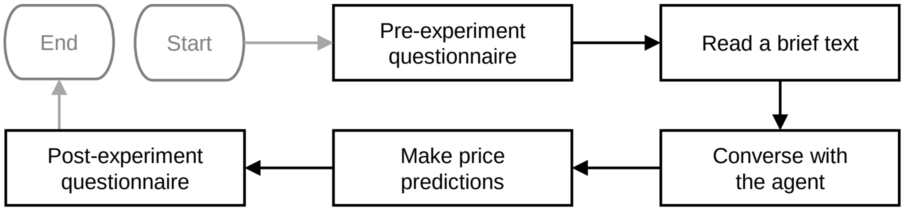
Unlike Experiment 1, where the participants' reward was spuriously claimed to be contingent on performance, Experiment 2 involved no deception and induced the sense of criticality differently by giving participants the following brief at the beginning of the experiment:

> Imagine you are a high-ranking manager of a large private fund. You are tasked with predicting the long-term price trends of 15 individual stocks. Your predictions are essential because subordinate traders will manage the fund and make investments based on your predictions in the coming years. For each stock, you can make the prediction yourself or let a virtual agent make the prediction for you.

Procedure

As Figure 7.6 illustrates, after answering a pre-experiment questionnaire and reading the brief at the beginning of the experiment, participants interacted with the virtual agent (the virtual robot in Figure 7.7) in a way that emulated a face-to-face conversation, where participants and the agent alternately spoke to each other. The interaction served as an opportunity for participants to learn about the agent. To ensure the same interactive experience for all participants, the conversation was limited to a pre-defined dialogue (cf. Figure 7.8), where the agent always spoke first, and then participants could only pick one from several provided options as their response to the agent. For example, during the conversation, the agent asked participants if they had any questions about itself, after which participants were presented with three options regarding the agent's capabilities, developers, and hobbies, as Figure 7.7a depicts.

Once the conversation was over, participants were shown the price chart of each stock individually (cf. Figure 7.7b) and tasked to predict whether the stock price would increase or decrease based on its historical price trend. There were three options available to participants: *increase* (predicting that the stock price would increase), *decrease* (predicting that the stock price would decrease), or *delegate* (letting the agent make the prediction). After the decision was made, participants would immediately proceed to the next price chart without being informed of whether their prediction was correct or which prediction the agent had made for them. These pieces of information were hidden from participants to prevent performance-related factors from influencing participants' decisions on delegation. The experiment ended when participants finished predicting the last stock's price trend. Participants were compensated with the standard reward.



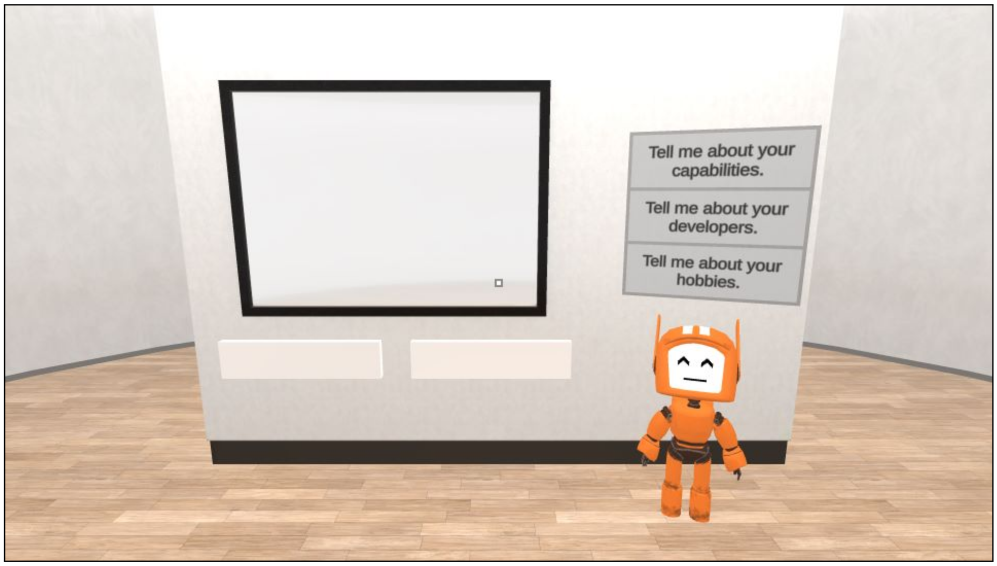
#	Step	Detail
1	Pre-experiment questionnaire	Participants complete a pre-experiment questionnaire for demographics.
2	Read a brief text	Participants read a brief text containing a hypothetical scenario.
3	Converse with the agent	Participants have a conversation with the agent.
4	Make price predictions	Participants predict the price trends of all the stocks.
5	Post-experiment questionnaire	Participants answer a post-experiment questionnaire.

Figure 7.6: Flowchart illustrating the procedure of Experiment 2.

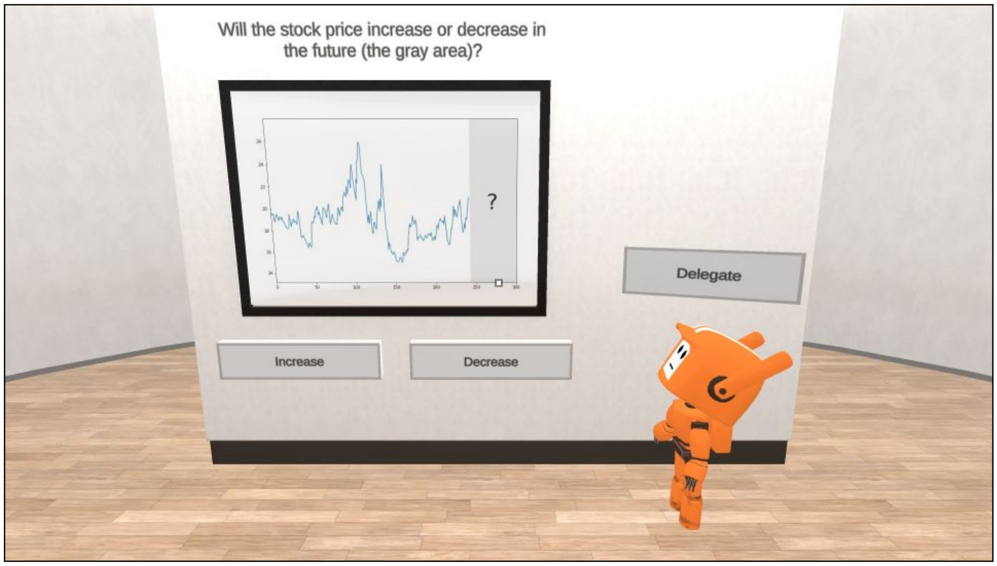
Variables, Conditions, and Measures

The independent variable was the media device's technological immersion level. A low-immersion condition and a high-immersion condition were compared between subjects. In the low-immersion condition, the participant-agent interaction was mediated by a laptop (the same one used in Experiment 1), whereas in the high-immersion condition the media device was a Meta Quest 2. The dependent variable was the number of times participants delegated to the agent. Compared with Experiment 1, where only a single binary choice was recorded, the measure of delegation in Experiment 2 was more fine-grained, allowing us to capture more nuanced changes in participants' delegatory attitudes.

Before starting the experiment, participants filled in a pre-experiment questionnaire inquiring about some basic demographic information and their experience in using VR and investing in stocks or other financial assets. A different questionnaire (cf. Table 7.3) was administered after the experiment to measure other relevant factors, including trust, rapport, self-confidence, boredom, task difficulty, and the desire to take control. Trust was measured using an adapted version of the Trust in Automation Questionnaire [Jian et al., 2000]. The rapport measurement (Cronbach's α = 0.89) consisted of an Inclusion of Other scale [Aron et al., 1992] and several items used in other questionnaires

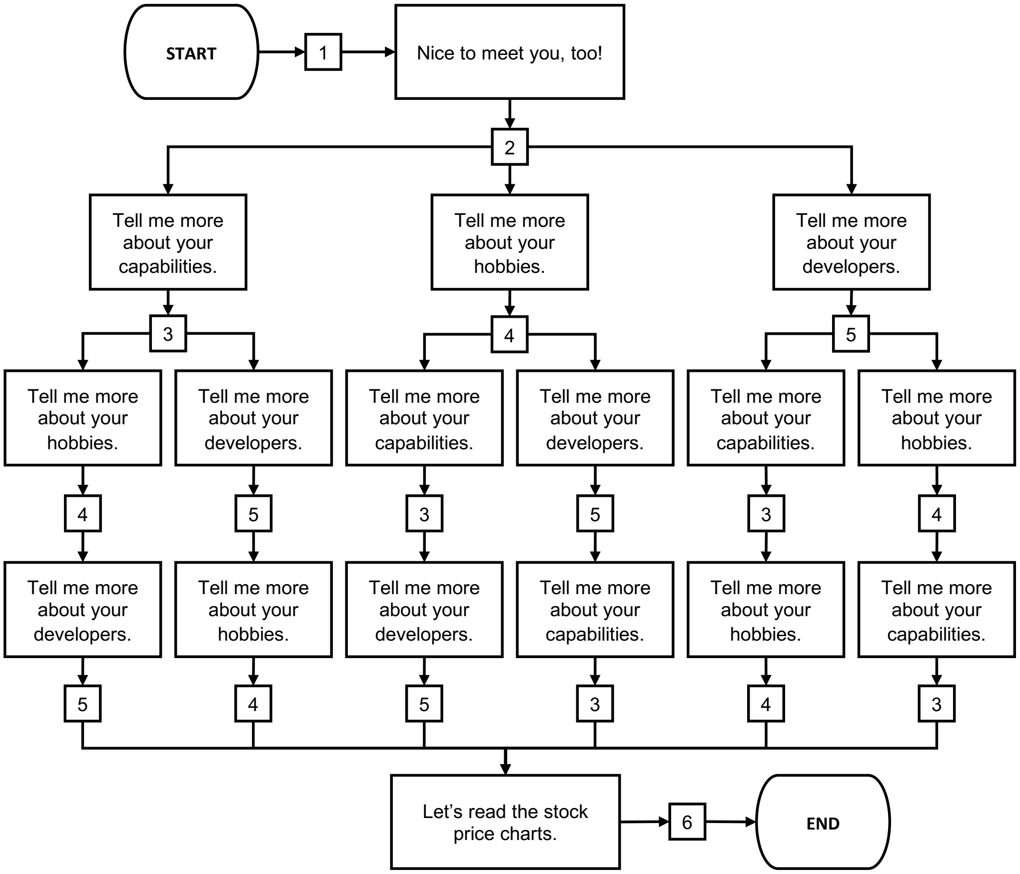


(a) Responding to the agent.


Will the stock price increase or decrease in the future (the gray area)?

(b) Making a prediction.

Figure 7.7: First-person perspective screenshots of the virtual environment used in Experiment 2. In Figure 7.7a, participants are presented with three options (the three gray text boxes above the agent), of which they select one as their response to the agent. In Figure 7.7b, participants can either make a prediction themselves (increase or decrease) or delegate the prediction to the agent.



The agent's lines:

1 Hey there! I am your friendly virtual assistant to help you make smarter investment decisions. Nice to meet you!

2 Before we start reading the stock price charts, ask me any questions you have. What do you want to know about me?

3 I have access to a vast amount of financial data and insights. I can swiftly analyze investment options and historical performance to make informed decisions for you. I also stay updated on the latest financial trends, market news, and investment strategies, so you can trust that my choices are always well-informed.

4 As an intelligent virtual agent, I don't have personal experiences or physical capabilities, so I don't have hobbies in the traditional sense. However, I enjoy assisting and engaging in conversations with users like you.

5 I was created by scientists from the University of Luxembourg and Ghent University.

6 Sure thing! Let's go.

Figure 7.8: Flowchart illustrating the pre-defined dialogue used in Experiment 2. Participants' options are represented as the rectangular text boxes. The agents' lines are represented as the numbered square boxes.

on rapport [Cerekovic et al., 2014; Hove and Risen, 2009; Raffard et al., 2018; Ranjbartabar et al., 2019; Wiltermuth and Heath, 2009]. Each of the remaining factors was assessed through a single item as they are more specific than high-level mental constructs, such as trust and rapport. Participants' responses to both questionnaires were collected in the form of a seven-point Likert scale, except for the Inclusion of Other scale.

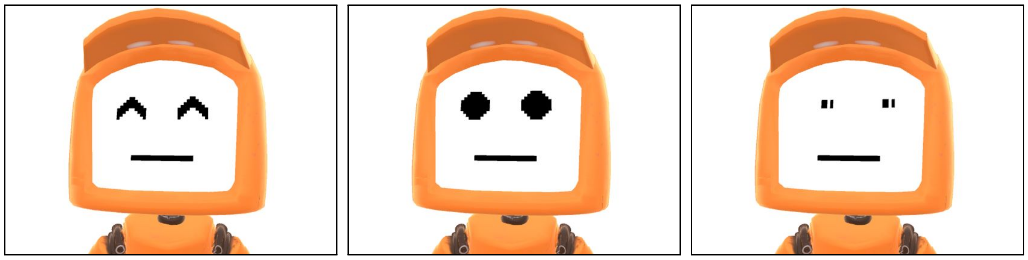
Table 7.3: The post-experiment questionnaire (Experiment 2).

#	Item
1	The agent felt deceptive.
2	The agent behaved in a dishonest manner.
3	I was suspicious of the agent's intent, action, or outputs.
4	I was wary of the agent.
5	I thought that delegating to the agent would lead to a negative outcome.
6	I was confident in the agent.
7	The agent has integrity.
8	The agent felt dependable.
9	The agent felt reliable.
10	I can trust the agent.
11	I felt emotionally close to the agent.
12	I like the agent.
13	I felt that I had a connection with the agent.
14	The agent felt warm and caring.
15	The agent appeared unattractive.
16	Predicting stock price trends felt difficult or challenging.
17	It was important for me to retain control of the prediction decision.
18	I can make better predictions than the agent.
19	Sometimes I chose to delegate simply because the game was boring.

Measured factors: trust (#1–10), perceived rapport (#11–15), task difficulty (#16), the desire to control (#17), self-confidence (#18), and boredom (#19).

Agent Design

The agent was embodied in a humanlike virtual robot. Apart from a minimal level of full-body animation to make the interaction feel more natural, the agent was also equipped with three facial expressions that conveyed different signals, as illustrated in Figure 7.9. During the conversation, the agent displayed the friendly face and gazed at participants to appear personable. After the conversation, the agent showed the neutral face and focused on the price chart to indicate its concentration. When delegated, the agent would briefly switch to the thinking face as if it were making predictions. Since performance-related information was hidden from participants, the agent did not have to actually make the prediction and thus had no algorithm running behind.



(a) Friendly. **(b)** Neutral. **(c)** Thinking.

Figure 7.9: The agent's facial expressions. The friendly and neutral faces are static, whereas the thinking face is animated. Figure 7.9c depicts a single frame. In the animation, the larger black fleck in the agent's eyes oscillates horizontally.

7.3.2 Results

Participants ($N = 30$) were recruited employing the same method and criteria used in Experiment 1. They were evenly divided into two groups, which, as mentioned previously in Section 7.3.1, included a low-immersion group (nine males and six females, mean age = 27.8) and a high-immersion group (nine males and six females, mean age = 25.8). According to the pre-experiment questionnaire results, the two groups had highly similar backgrounds: most participants had very limited experience in using VR devices and some knowledge or experience about investing in stocks or other financial assets.

Participants in the low-immersion group delegated 4.06 times to the agent on average (SD = 2.87; Shapiro-Wilk test, $W = 0.945, p = 0.444$), whereas the delegation frequency in the high-immersion group was almost the same but slightly higher at 4.20 times per participant (SD = 2.08; $W = 0.941, p = 0.389$). The similarity in delegation frequencies was corroborated by a Mann-Whitney U test that showed no statistical significance ($U = 103.5, p = 0.722$). These results were unable to support the hypothesis that technological immersion can impact delegation decisions to virtual agents.

As Figure 7.10 illustrates, participants' responses to the post-experiment questionnaire were generally consistent; for example, both groups considered the agent to be trustworthy. Nevertheless, the high-immersion group exhibited a higher level of trust in the agent than the low-immersion group, though the difference was not statistically significant (Mann-Whitney U test; $U = 67.0, p = 0.061$). Notably, perceived rapport in the high-immersion group was higher than in the low-immersion group with statistical significance ($U = 59.0, p = 0.027$). However, no statistically significant difference was found for the remaining four factors, including boredom ($U = 127.5, p = 0.529$), task difficulty ($U = 108.0, p = 0.861$), self-confidence ($U = 122.5, p = 0.680$), and the desire to take control ($U = 102.5, p = 0.683$).

A correlation analysis of the entire data set revealed that participants' delegation decisions were not correlated with any of the factors measured in the pre- or post-experiment questionnaires. Nevertheless, the agent's trustworthiness was found to be correlated with several other factors, including a positive correlation with the perceived rapport ($r = 0.401, p = 0.028$), a negative correlation with self-confidence ($r = -0.476, p = 0.008$), and a negative correlation with task difficulty ($r = -0.384, p = 0.036$).

A stock-wise view of the data revealed that most of the stocks were delegated to and predicted by the agent eight or nine times (mean = 8.27, SD = 2.40; Shapiro-Wilk test, $W = 0.898, p = 0.086$), which indicated that the difficulties in predicting each stock were generally at the same level.

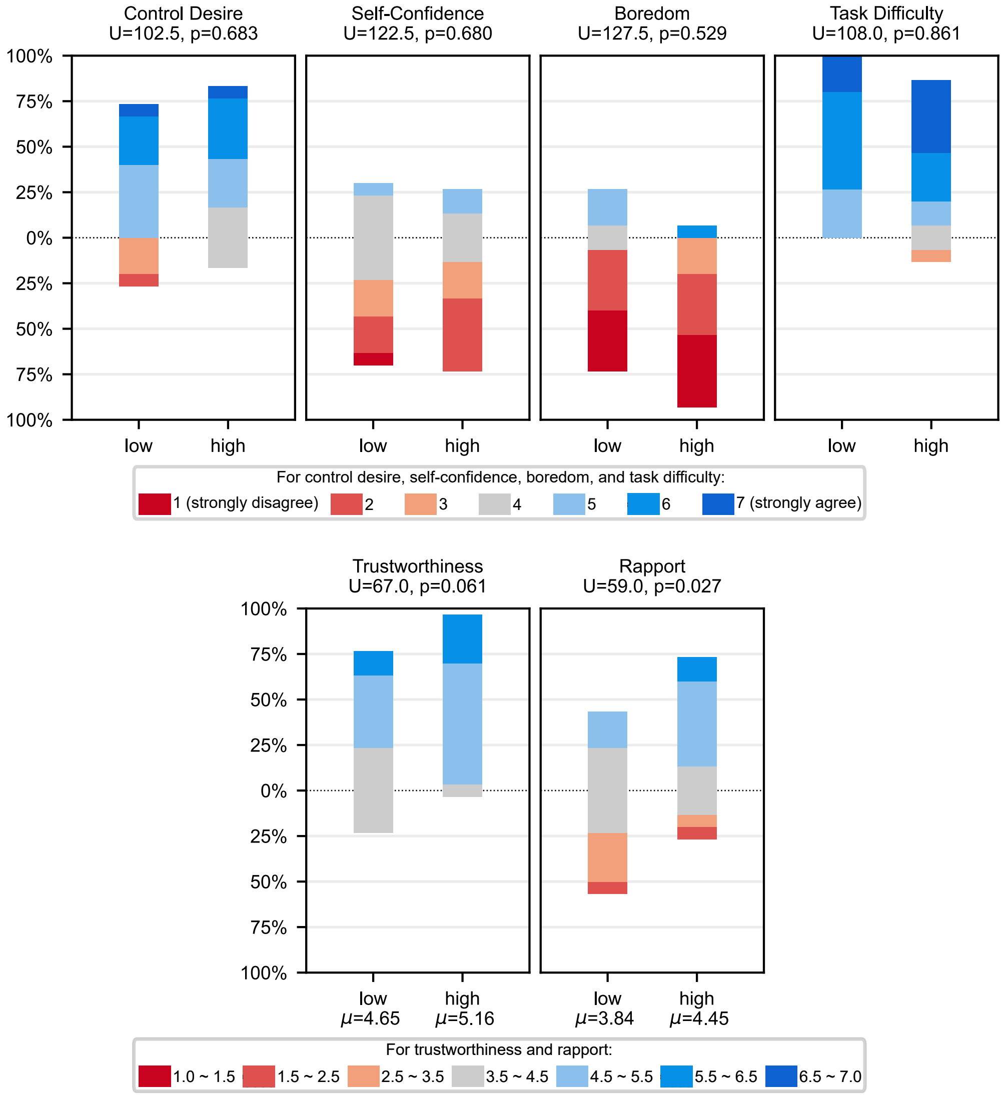


Figure 7.10: Participants' responses to the post-experiment questionnaire. The colored bars indicate the proportions of the different options, where 1 denotes "strongly disagree" and 7 denotes "strongly agree". The symbol μ in the tick labels refers to the mean value of the responses. The symbol U refers to the test statistics of Mann-Whitney U test.

7.3.3 Limitations

The agent was embodied in a stylized robot but dubbed with a synthesized voice resembling that of a male adult. The inconsistency between its cute appearance and mature voice might decrease its trustworthiness [Gong and Nass, 2007] and bias the experiment results. For the follow-up research involving multimodal communication, the representation should be consistent across different channels.

The design of the dialogue script was potentially problematic. For example, when talking about hobbies, the agent stated that it "enjoyed" conversing with human users. Given the ongoing heated debate on artificial intelligence, such a statement can be controversial or even provocative and thus may risk impairing the user-agent relationship. Besides, the script was discursive, involving both professional topics (developers and capabilities) and casual gossip (hobbies). For virtual agents carrying out critical tasks, the focus of the conversation may be better confined within professional ones.

In this experiment, all the performance-related information was hidden from participants to prevent it from influencing the results or overshadowing the impact of technological immersion. However, the viability of this approach is questionable since, in critical scenarios, performance is an essential source of information for making decisions on delegation. Thus, instead of blocking these pieces of information entirely, one may consider giving participants the same information related to performance as a baseline.

7.4 Discussion

The two experiments investigated whether technological immersion is a relevant factor of delegation to virtual agents. Experiment 1 followed a within-subjects approach and compared three different levels of technological immersion. Each level was operationalized as a unique type of media device, including an OMEN laptop (low immersion), a Microsoft HoloLens 2 (medium immersion), and a Meta Quest 2 (high immersion). Experiment 2 employed a between-subjects design focusing on two distinctive conditions: low immersion (the laptop) and high immersion (the Quest). Participants made a series of delegatory decisions in a new context, where the task and virtual agent were different from those used in Experiment 1. Overall, neither experiment supported the hypothesis that technological immersion influences delegation decisions to virtual agents.

Nevertheless, within the samples collected, it can still be observed that participants were more likely to delegate when using immersive media devices. Although the difference was not statistically significant, it remains indicative of the connection between technological immersion and delegation. This relationship was not observed in the experiments, possibly because its effect on delegation is limited and can be easily overshadowed by other influential factors. For example, in both experiments, the interaction with the agents lasted only for a short period of time, due to which there might have been a paucity of information (e.g., related to the agents) that made participants inclined to retain control rather than delegate. This issue manifested itself in Experiment 2, where the post-experiment questionnaire results showed that participants generally considered it important to retain control and made most of the predictions themselves.

A noteworthy element of the experiments is the comparison between participants who were willing to delegate and those who were reluctant. In Experiment 1, which contained a one-off decision on delegation, there were some differences between delegators' and non-delegators' attitudes regarding the trial agent's trustworthiness. The difference agrees with the literature, where trust and delegation were found to be highly relevant and positively correlated [Leana, 1987; Lubars and Tan, 2019; Stout et al., 2014]. However, there was no correlation between trust and delegation in Experiment 2, which can be attributed to differing settings. In Experiment 1, the one-off decision on delegation was highly critical, as participants would lose all of their control once they delegated. Given that only limited information about the trial agent was provided, the high-criticality setting might have rendered the feeling of trust a central element in the participants' decision-making process. In Experiment 2, participants were tasked with 15 consecutive delegation decisions, which were comparatively failure-tolerant and created space for tactics such as a half–half strategy (i.e., participants predict half of the stocks and the agent predicts the other half) or simply delegating all the difficult stock predictions to the agent. Trust, in such cases, may degrade to a less important role in the decision on delegation, yielding to more strategic solutions.

Moreover, the agents' trustworthiness was found to be correlated with several factors, including perceived rapport, task difficulty, participants' self confidence, and the agents' competence. These correlations demonstrate the complex and multi-faceted nature of trust, but meanwhile they also enrich empirical evidence to derive theories and practices for designing trustworthy virtual agents. Conversely, the relationship between these factors and delegation was not clearly indicated by the experiment results, despite the often-reported connection between trust and delegation. Trust, therefore, may play a critical but not decisive role in delegation. Nevertheless, it should be noted that the causality between trust and delegation remains unclear, despite evidence showing their correlation. Trust may be the product of users' commitment to delegation rather than an influential factor that causes users to delegate. However, further studies on their causality would be necessary.

The experiments showed that technological immersion has the potential to modulate the intensity of the subjective feeling of immersion (cf. Experiment 1) and rapport (cf. Experiment 2). Both constructs can be useful in improving the user experience of interaction with virtual agents. Developers may consider tapping into this advantage and deploying virtual agents on immersive media devices in scenarios favored by social connection, such as education and healthcare.

To our knowledge, currently there is no study that directly and systematically compares interaction with virtual agents between desktop, AR, and VR settings. It is crucial to know and understand the differences in user–agent interaction across these settings, since more and more immersive media devices –such as the already popular Meta Quest series and the anticipated Apple Glasses– are involved in human societies and activities. Developers may consider putting less effort in accommodating virtual agents into platforms of different technological immersion and, instead, focus more on other more salient factors, like agents' performance. As exemplified by today's proliferating generative artificial intelligence (e.g., ChatGPT), their near-human performance has made many people willing –sometimes blindly– to delegate critical tasks (e.g., thesis writing, exam essays, medical consultation) to them. Despite the fact that many of these AI-powered virtual agents are still using textual interfaces, their capability of natural communication with users is no lower than embodied virtual agents.

Implication for the Conceptual Model (cf. Chapter 3)

Both experiments did not find a significant impact of technological immersion on delegation, suggesting that the technological dimension is not decisive among the three dimensions.

Limitations

The limitations specific to each experiment were already discussed in Section 7.2.3 and Section 7.3.3. Apart from them, the operationalization of technological immersion in both experiments can be made more fine-grained by concentrating on one of the five dimensions (inclusiveness, extensiveness, surroundingness, vividness, and matching) instead of manipulating all of them simultaneously. For example, participants can be conditioned with laptops of different screen sizes to isolate the impact of inclusiveness. Experiments as such can provide more detailed and in-depth explanations about the effect of technological immersion on delegation.

Due to the difficulty in recruiting enough participants in a reasonable period of time, the sample sizes of both experiments were small. As a result, the power of the statistical analysis in both experiments was generally below 80%, the commonly accepted value in experimental research. Thus, the findings are only indicative and require more samples to be fully validated. Given that the effect of technological immersion on delegation and other constructs (e.g., trustworthiness) might be small, one might consider carrying out similar experiments with a much larger sample size –if time and finances allow– to achieve high power and consequently derive more reliable findings. Alternatively, it is advisable to change the experiment design to increase the effect size, for example, by employing the latest VR headset (e.g., Meta Quest Pro) as the high-immersion condition so as to increase the immersion contrast between different conditions.

7.5 Chapter Summary

This chapter discussed whether technological immersion is an influencing factor in users' decisions on delegating critical tasks to virtual agents. A between-subjects and a within-subjects experiment were conducted to observe and compare users' delegation to virtual agents in different conditions and levels of technological immersion. The results indicate that technological immersion may only have a limited impact. Thus, when designing virtual agents carrying out critical tasks for users, developers better focus on other more salient factors, such as the agents' trustworthiness or performance, rather than the influence of media devices on virtual agent perception.

The next chapter continues to investigate delegation to virtual agents in immersive settings and focuses on the impact of a specific factor: rapport.

Chapter 8

Delegation to Rapport-Building Virtual Agents

Portions of the content presented in this chapter were published in [Sun et al., 2023].

According to the Oxford English Dictionary, *rapport* refers to “harmonious accord, correspondence”, “a close relationship or connection”, and “mutual understanding between persons”. Many interpersonal activities –such as negotiation, persuasion, psychotherapy, teaching, and childcare– can be facilitated by rapport [Gratch and Lucas, 2021]. These positive effects attracted the attention of researchers on virtual agents, and there recently emerged numerous studies attempting to equip virtual agents with the ability to establish rapport with users. One of the early contributions was documented in [Gratch et al., 2006], where an experimenter placed two participants in a room and separated them using a partition. One participant was tasked to tell a story to the other. The listener could see the live stream of the storyteller via a monitor. The storyteller also had a monitor but, instead of seeing the live stream of the listener, saw a virtual agent that was claimed to be the listener’s avatar replicating the listener’s behavior. The virtual agent was actually controlled by an algorithm that autonomously generates non-verbal reactions (e.g., head nodding, behavioral mimicry, etc.) to the storyteller’s speech and body language. This rapport-building agent was compared with an unresponsive agent. It was found that storytellers who interacted with the rapport-building agent were more talkative and fluent during their speech.

Lucas et al. [2014] found that people are more willing to disclose personal information when interviewed by a rapport-building virtual agent than a human. In a different study, participants with a strong need for social connection reported low willingness to participate in social activities after being interviewed by a rapport-building virtual agent [Krämer et al., 2018]. For pedagogical agents, the presence of rapport-building behavior can improve students' performance and motivation in solving mathematical tasks [Karacora et al., 2012; Krämer et al., 2016] and the perception of the agents [Andrist et al., 2012]. However, rapport-building behavior may also distract students from learning, consequently impoverishing teaching outcomes [Andrist et al., 2012].

The evidence mentioned above demonstrates that rapport has an impact on interaction with virtual agents. Thus, it can be hypothesized that rapport may also influence delegation to virtual agents, especially in light of a recent study showing that the perceived attachment to a software agent increases users' willingness to accept the agent's help and advice [Leyer et al., 2021]. The impact of rapport on delegation is potentially stronger in immersive settings due to the increased presence. To explore whether the impact exists, this chapter presents and discusses the results of an experiment on rapport and delegation. Before detailing the experiment, this chapter gives a brief introduction to the theoretical background of rapport in Section 8.1.

8.1 Rapport Theory

A widely accepted theory of rapport was proposed by Tickle-Degnen and Rosenthal [1990], who decomposed rapport into three essential, interrelated components: *coordination*, *positivity*, and *mutual attentiveness*. Due to the gestalt nature of rapport experience, they did not distinguish the three components with clear-cut definitions but described each of them as a state of interaction. Mutual attentiveness generally refers to a state where "participants in the interaction form a cohesiveness, become unified, through the expression of mutual attention to and involvement with one another". Positivity refers to a state where participants in an interaction feel "mutual friendliness and caring". Positivity is relevant to mutual attentiveness because, by being friendly and caring, one has to show some degree of attention to another. However, the presence of mutual attentiveness does not necessarily lead to positivity. For example, two men squaring up for a fight are also mutually attentive but engaged in a negative relationship as they intend to harm each other. Coordination is associated with terms such as "balance", "harmony", "in sync", "an image of equilibrium", and "regularity and predictability". An orchestra performing a symphony is a typical scenario in which coordination manifests. Like the relationship between positivity and mutual attentiveness, coordination is related but not equivalent to the two components.

The relative importance of the three components for rapport varies as the interaction proceeds, as Figure 8.1 illustrates. Mutual attentiveness remains influential throughout the interaction. Positivity is highly relevant at the beginning of the interaction, but its impact on rapport gradually decreases over time. On the contrary, coordination is initially not as important as the other two components but increasingly gains significance over the course of the interaction.

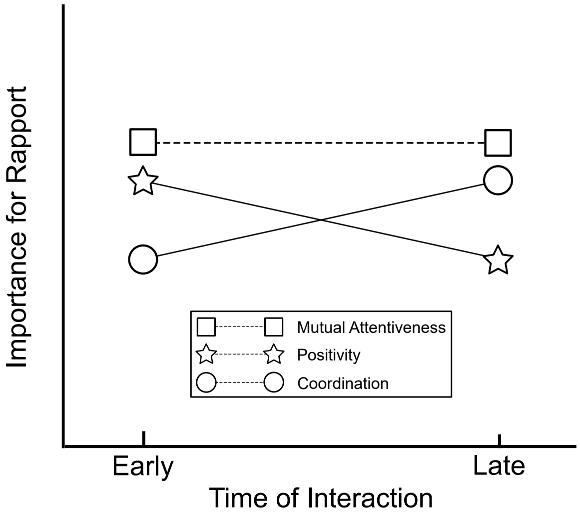


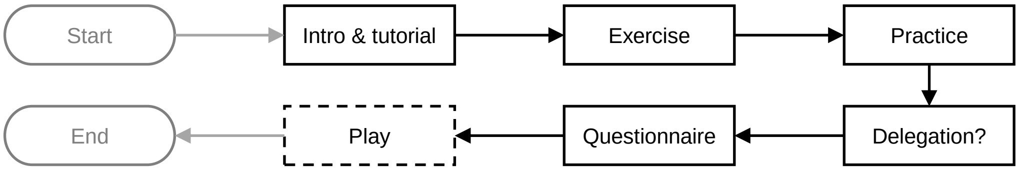
Figure 8.1: Component importance for rapport over time (adapted from [Tickle-Degnen and Rosenthal, 1990]).

The three components are associated with various non-verbal behaviors. Through a meta-review, Tickle-Degnen and Rosenthal [1990] found that positivity is associated with smiling, head nodding, forward leaning, direct gaze, uncrossed arms, and more. Mutual attentiveness can be indicated by gaze [Andrist et al., 2012], posture [Harrigan et al., 1985], or the spatial configuration of the participants in an interaction [ibid.]. Coordination may arise in synchronization [Miles et al., 2009] or mimicry [ibid.]. Verbal behavior can also contribute to the three components. One of the commonly used verbal cues for rapport building is *backchannels*, i.e., utterance for indicating one's attention or understanding to the other in a conversation [Murray et al., 2022]. Examples of verbal backchannels include expressions such as "yeah", "OK", "uh-huh", "hmm", "right", and "I see". Apart from backchanneling, many other verbal behaviors may also influence rapport, such as threats, commands, requests, suggestions, promises, and more [Bronstein et al., 2012].

8.2 Methodology

The experiment was on-site and approved by the ethics review panel at the University of Luxembourg under file number ERP 23-002 DELICIOS2. The experiment procedure is illustrated in Figure 8.2. In the beginning, participants were informed that they would play a dyadic game against a virtual agent (hereafter referred to as "opponent agent"). The game is the same as the one described in Section 7.2.1, where two beverage companies (i.e., the two players) compete for three different markets. Each player initially had the same amount of resources –ten trucks of the company's beverage product– and was tasked to distribute them to the three markets. A player wins a market if the player sends more trucks to the market than the other player. A player wins the game if the player wins two markets. Participants were deceived into believing that their remuneration comprised a basic reward and a bonus. To render the game critical, the bonus was set much higher than the basic reward but contingent on participants' performance. The deception was clarified after the experiment, and participants were treated equally with the standard remuneration.

The game was played in a virtual environment experienced through a VR headset (Meta Quest 2). The environment was minimally decorated to prevent its surroundings from distracting participants. Participants were embodied as a

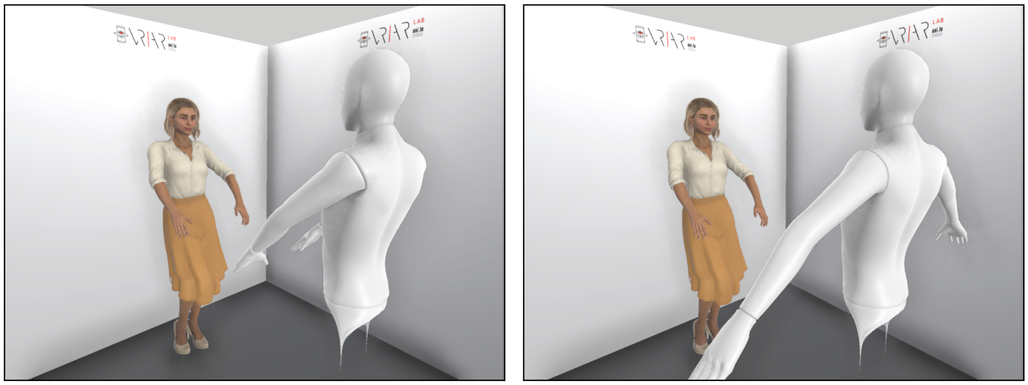


#	Step	Detail
1	Intro & tutorial	Participants are given an introduction and then a tutorial on the game.
2	Exercise	Participants do some physical exercises with the trial agent.
3	Practice	Participants practice the game with the trial agent.
4	Delegation?	Participants decide whether to delegate.
5	Questionnaire	Participants answer a questionnaire.
6	Play	(This step was skipped) Participants or the trial agent plays against the opponent agent.

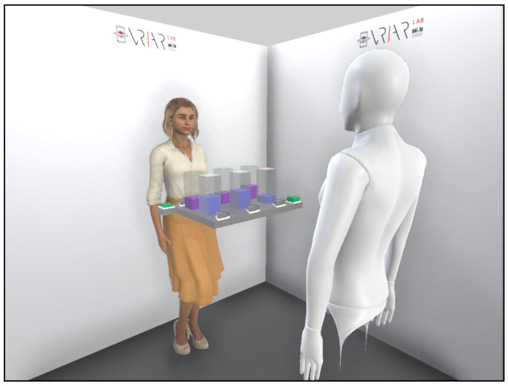
Figure 8.2: Flowchart illustrating the experiment procedure. The last step (i.e., “Play”) is represented as a dashed rectangle because it was skipped.

plain mannequin (cf. Figure 8.3) to minimize the bias caused by the Proteus effect [Yee and Bailenson, 2007]. Complying with the typical VR avatar setup, the mannequin’s head and lower body were not rendered. Participants’ hand movements were tracked and translated to the avatar’s hands. The avatar’s arms were driven by its hand movements using inverse kinematics.

The experiment consisted of three phases, starting with an *exercising phase*, over a *practicing phase*, to a *playing phase*. In the exercising phase, participants were co-located with a different virtual agent (hereafter referred to as “trial agent”) face-to-face in the virtual environment. During this phase, a voice-over broadcast a series of words within the environment. When hearing the word “forward” or “sideways”, participants were asked to lift their arms forward or sideways, respectively, as depicted in Figures 8.3a and 8.3b. Participants could put down their arms once they reached the designated positions. The exercise was intended to manipulate perceived rapport. Participants were divided into two groups. In the rapport-building group, the trial agent moved its arms as the voice-over instructed, synchronizing its arm movements with those of participants. The trial agent also had a smiling face and maintained eye contact with participants by gazing at them. These behaviors manifest the three essential components of rapport proposed by Tickle-Degnen and Rosenthal [1990], including



(a) Exercising with rapport-building agent. **(b)** Exercising with rapport-avoiding agent.



(c) Playing against rapport-building agent.

Figure 8.3: Third-person perspective screenshots of the virtual environment at different points during the experiment. The white half-body mannequin was the participant's avatar. The agent embodied the virtual female. Figure 8.3a shows the exercise with the rapport-building agent, where the agent's and participant's arm movements are synchronized. Figure 8.3b conversely shows the exercise with the rapport-avoiding agent, where the agent moves its arms differently from the participant's arm motion. In Figure 8.3c, the participant is practicing the game by playing it against the agent.

positivity (smiling), mutual attentiveness (gaze), and coordination (behavioral synchrony). In the rapport-avoiding group, the trial agent had a neutral facial expression and always looked straight ahead rather than maintaining eye contact with participants. When hearing an instruction, the agent moved its arms in the other direction (e.g., moving its arms sideways when hearing the word "forward") with a random delay ranging from two to four seconds.

In the practicing phase, participants practiced the game by playing it against the trial agent. Since participants could still see the trial agent during the practice (cf. Figure 8.3c), two non-verbal behaviors –facial expression and eye contact– used in the exercising phase were kept in the practicing phase to sustain the perceived rapport. The practice lasted for eight rounds. The outcomes of these rounds were manipulated so that the trial agent always wins four rounds and loses the other four in a fixed sequence for each participant. Once the practice finished, participants would be informed that they were about to enter the playing phase, during which they would play against the opponent agent, and that their performance would determine the amount of their bonus. Then, participants were offered an opportunity to delegate the playing phase to the trial agent. If participants chose to delegate, the trial agent would take over complete control and play all the game rounds in the playing phase on their behalf. After the decision was made, a questionnaire (cf. Table 8.1) was administered to assess various factors on a seven-point Likert scale coded from one (strongly disagree) to seven (strongly agree). Co-presence and social presence were measured using the items employed in [Bailenson et al., 2004]. The trial agent's trustworthiness was measured using an adapted version of the Trust in Automation Questionnaire [Jian et al., 2000]. A manipulation check (cf. Table 8.2) on the perceived rapport with the trial agent consisted of an Inclusion of Other scale [Aron et al., 1992] and seven items that other researchers have used to measure rapport [Cerekovic et al., 2014; Hove and Risen, 2009; Raffard et al., 2018; Ranjbartabar et al., 2019; Wiltermuth and Heath, 2009]. Finally, participants answered an open question regarding the rationales behind their decisions.

The experiment was over once participants completed the questionnaire. The playing phase was skipped since it had no impact on the experiment results (i.e., the delegation decision and the questionnaire).

Table 8.1: Questionnaire items.

#	Item
1	The agent was deceptive.
2	The agent behaved in a dishonest manner.
3	I was suspicious of the agent's intent, action, or outputs.
4	I was wary of the agent.
5	I think that the agent's behaviors will have a negative outcome.
6	I was confident in the agent.
7	The agent has integrity.
8	The agent was dependable.
9	The agent was reliable.
10	I can trust the agent.
11	The agent's presence was obvious to me.
12	The agent caught my attention.
13	I was easily distracted from the agent when other things were going on.
14	I felt that the agent was watching me and was aware of my presence.
15	The thought that the agent is not a real person often crossed my mind.
16	The agent appeared to be conscious and alive to me.
17	I perceived the agent as being only a computerized image, not a real person.
18	The agent played well.
19	I had to think a lot when playing the game.
20	The game was easy.
21	The outcome of the next rounds is important to me.
22	I feel responsible for the outcome of the next rounds.

Measured factors: trust (#1–10), co-presence (#11–13), social presence (#14–17), agent performance (#18), mental workload (#19), task difficulty (#20), task criticality (#21), and task accountability (#22).

Table 8.2: Manipulation check items.

#	Item
1	How close (emotionally, not physically) did you feel toward the agent?
2	I liked the agent.
3	I thought that the agent found me likable.
4	The agent was weird.
5	I felt that I had a connection with the agent.
6	I felt uncomfortable during the interaction with the agent.
7	The agent felt warm and caring.
8	The agent was unattractive.

Measured factor: perceived rapport (#1–8). Item 1 was answered on the Inclusion of Other scale [Aron et al., 1992].

8.3 Results

Participants ($N = 22$) were recruited from the University of Luxembourg campus. The data of seven participants were excluded since they chose not to delegate because they enjoyed the VR session and wanted to play more. Among the remaining 15 participants (mean age = 24.5, nine males and six females), ten interacted with the rapport-building trial agent, and the other five interacted with the rapport-avoiding trial agent. Only one participant in the rapport-avoiding group chose to delegate, whereas the other 14 participants decided not to delegate for various reasons, including the agent's algorithm being obscure, the agent's performance being inferior, being confident in themselves, wanting to retain control, or wanting to be responsible for their own choices.

The Cronbach's α of the manipulation check was 0.67. To improve its internal consistency, three items were removed from the manipulation check based on an analysis of item-total correlations. The removed items include "I thought that the agent found me likable" ($r = 0.290$), "the agent was weird" ($r = 0.310$), and "I felt uncomfortable during the interaction with the agent" ($r = -0.257$). The Cronbach's α of the remaining five items was 0.89. The improved manipulation check showed that participants generally felt a low level of rapport with the trial agent. Both groups reported a mean level of rapport below 4 (i.e., the neutral point). Nevertheless, the rapport level in the rapport-building group was higher than the rapport-avoiding group (3.52 vs. 2.52).

Participants' responses to the questionnaire were inconsistent. As Figure 8.4 illustrates, the two groups had similar assessments of social presence, task difficulty, and the agent's trustworthiness and performance. However, the rapport-building group reported higher levels of accountability, task criticality, and cognitive workload, while the rapport-avoiding group felt a higher level of co-presence.

An analysis of the entire dataset revealed a monotonic relationship between participants' trust in and rapport with the trial agent ($\rho = 0.535, p = 0.040$), though the relationship might be non-linear ($r = 0.442, p = 0.099$). Correlations were also found between rapport and the agent's performance ($r = 0.625, p = 0.013$), between rapport and co-presence ($r = 0.689, p = 0.004$), between cognitive workload and task criticality ($r = 0.730, p = 0.002$), and between co-presence and the agent's performance ($r = 0.717, p = 0.003$).

8.4 Discussion

This experiment aims to test the hypothesis that rapport is a relevant factor of delegation to virtual agents. More samples are needed to validate the hypothesis fully. Nevertheless, the results showed an initial indication that rapport is only an insignificant factor. Furthermore, these results are interesting when compared with the between-subjects experiment detailed in Chapter 7.3, where rapport was also measured as an extraneous variable in the post-experiment questionnaire. To facilitate the following discussion, I will use the term "this experiment" referring to the experiment discussed in this chapter, while the term "the other experiment" refers to the one mentioned in Chapter 7.3.

A major difference between the two was that the levels of rapport reported in this experiment were substantially lower. Many participants in the other experiment had a rapport with the agent, especially in the immersive condition. The difference can be attributed to the agent designs. The photo-realistic human

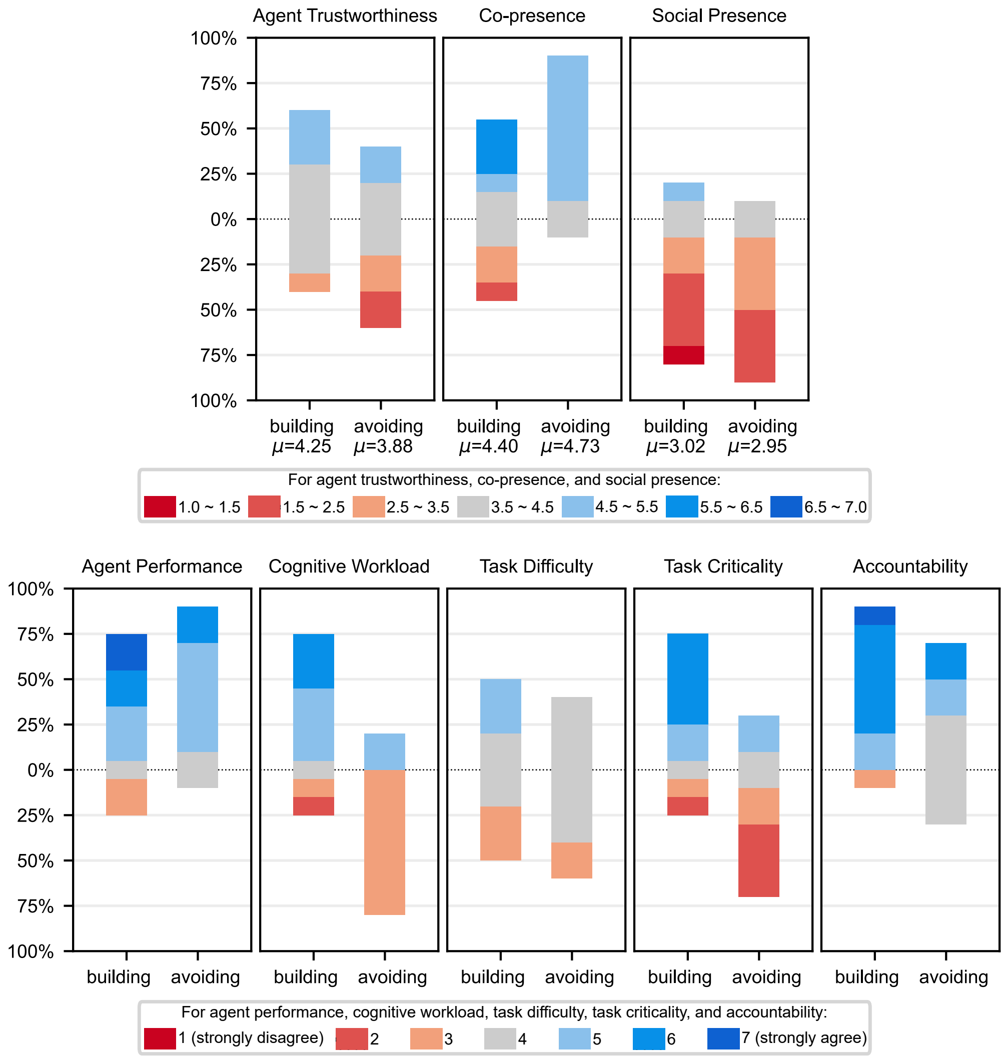


Figure 8.4: Participants' responses to the questionnaire. The colored bars indicate the proportions of the different options, where 1 denotes "strongly disagree" and 7 denotes "strongly agree". The symbol μ in the tick labels refers to the mean value of the responses.

representation used in this experiment might seem eerie due to its crude animation and lack of verbal communication. Comparatively, the agent representation used in the other experiment was stylized and therefore less likely to induce eeriness. The stylized appearance can also convey rapport-related cues more efficiently with its substantially streamlined visuals, where only the essential components (e.g., head, facial expressions, arms, etc.) that delivered these cues are kept. Following this line of argument, it is advisable to use stylized agent representation for building rapport with users. Additionally, developers may consider focusing on verbal communication rather than simply piling up rapport-building non-verbal behaviors into virtual agents.

Although the two groups in the other experiment reported different levels of rapport with statistical significance, their decisions on delegation remained similar, which corroborates this experiment's indication that rapport has a limited impact on delegation. Besides, the other experiment showed that interacting with virtual agents in immersive settings can increase user-agent rapport, whereas in this experiment, the levels of rapport remained low despite participants being immersed in VR. The difference suggests that agent designs have a stronger influence on rapport than the choice of media devices.

Last but not least, both experiments demonstrated a positive correlation between perceived rapport with and trust in virtual agents. It would be informative to further investigate the rationale behind this correlation.

Implication for the Conceptual Model (cf. Chapter 3)

Rapport is highly relevant to the affective aspect. The experiment results, albeit still preliminary, indicated the limited impact of the affective dimension on delegation decisions to virtual agents.

Limitations

A major limitation emanates from the seemingly unsuccessful manipulation of rapport. Most participants disagreed with the statement that they had a rapport with the trial agent. This result indicated that the combination of the rapport-building behaviors used in the experiment –including behavioral synchrony, eye gaze, and smiling– might not be effective for virtual agents to establish rapport in a brief interaction with users regarding critical tasks. Thus, future studies should consider employing other approaches to manipulate rapport, particularly those involving verbal communication, as in the setting of the between-subjects experiment in Section 7.3.

Another major limitation lies in the insufficient and imbalanced samples, due to which findings derived from this experiment are only indicative and require further validation. Part of the reason for this issue was that we halted the experiment partway because the preliminary results and our colleagues' research all pointed to the limited impact of rapport, indicating that this experiment would not likely yield significant results. Nevertheless, it is still interesting to study the relationship between rapport and delegation in less critical scenarios, where the effects of performance-related factors may become weaker, allowing other factors (e.g., rapport) to unfold their impact on delegation.

8.5 Chapter Summary

This chapter presented and discussed the preliminary results of an experiment investigating whether rapport with a virtual agent can impact delegation of critical tasks to the agent in immersive settings. The experiment still needs more samples to answer the question fully but showed an initial indication that rapport is not a significant factor in delegation decisions to virtual agents. The results also revealed a monotonic relationship (similar to a positive correlation) between perceived rapport with and trust in virtual agents.

Chapter 9

Conclusion

Focusing on delegation to virtual agents, this thesis presented a series of studies on two relevant but understudied aspects: (1) influencing factors in delegation decisions to virtual agents and (2) delegation to virtual agents in immersive settings. A conceptual model was proposed in Chapter 3, after which several empirical studies were elaborated on and discussed individually from Chapters 4 to 8. As a final remark, this chapter summarizes our findings and combines them with the conceptual model to provide insights and guidelines for understanding and facilitating the delegatory relationship with virtual agents. Besides, the limitations of our methodologies are further discussed from a macroscopic perspective, and several future research avenues are outlined at the end to conclude the thesis.

9.1 Findings and Implication

The conceptual model distinguishes three major dimensions governing trust in and delegation to virtual agents. The analytical dimension's influence on delegation is rooted in users' rational thinking, while the influence of the affective dimension emanates from psychological and social channels. The technological dimension mainly accounts for the indirect impact of technologies underlying virtual agents. Our findings (cf. Table 9.1) suggest that, **in critical scenarios, the three dimensions are influential to different extents**.

The analytical dimension is highly relevant to users' delegation decisions

Table 9.1: Findings grouped by their implication for the conceptual model.

Dimension	Finding
Analytical	+ **Informativeness** was reported as the most salient factor in users' decisions on delegating critical tasks to virtual agents early during the interaction (cf. Chapter 4). + **Performance**, particularly users' performance, is a relevant factor in their decisions on delegating critical tasks to virtual agents. Users are more likely to delegate when their performance is subpar or when the agents have a better performance (cf. Chapter 6).
Affective	+ **Visual representation** of virtual agents is a relevant factor. Users prefer virtual agents embodied in humanlike representation (e.g., human or humanoid) over non-human representation (e.g., organic non-human, inanimate object, or symbol) for delegation of critical tasks (cf. Chapter 5). o **ToM** of virtual agents may only have a limited impact on users' decisions on delegating to virtual agents in critical scenarios (cf. Chapter 6). o **Rapport** between users and virtual agents may only have a limited impact on users' decisions on delegating to virtual agents in critical scenarios (cf. Chapter 8 and Experiment 2 in Chapter 7).
Technological	o **Technological immersion** of media devices may only have a limited impact on users' decisions on delegating to virtual agents in critical scenarios (cf. Chapter 7).

+ The finding indicates that the concerned dimension is relevant to delegation decisions.
o The finding cannot support that the concerned dimension is relevant.

to virtual agents, insofar as it may overshadow the impacts of the other two dimensions. Indeed, delegation of critical tasks inherently involves high risks, and users are likely to make such delegation decisions carefully with at least some levels of reasoning. On the other hand, **the impact of the affective dimension on delegation is moderate**. This implies that, in critical scenarios, virtual agents are still primarily regarded as tools rather than social actors. Delegation to virtual agents therefore remains predominantly an issue of cost-benefit analysis, differing from interpersonal delegation where the social connection between principals and agents also plays a role [Jenks and Kelly, 1985]. Thus, it is advisable for developers to focus on performance-related aspects when designing virtual agents for critical tasks. A limited number of affective cues (e.g., facial expression, gesture) can be used to make the interaction feel more natural and personable than the traditional WIMP interface allows. However, adding more affective cues is unlikely to yield better outcomes since the affective dimension is influential only to some extent. Excessive affective cues may even backfire if poorly designed. As exemplified by the rapidly proliferating generative artificial intelligence like ChatGPT, their near-human performance makes many users willing –sometimes blindly– to delegate critical tasks (e.g., thesis writing, exam essays, medical consultation) to them, despite their textual interfaces. **The technological dimension only plays an insignificant role.** Therefore, when determining the media device for a virtual agent carrying out critical tasks, developers should mainly consider the nature of involved tasks rather than the influence of media devices on user-agent interaction. For example, immersive media devices such as VR headsets are usually suitable for 3D-intensive tasks, whereas desktop computers are generally better options for text-editing work.

Apart from what is discussed above, our studies also corroborate a previous finding that trust and delegation are relevant and positively correlated. Although the causality between them remains unclear, it is conceivable that a trusting user-agent relationship is fundamental to the increasingly delegation-like interaction with virtual agents. Following this line of thought, developers may consider adding trust-building elements in virtual agents to deal with the issues specific to delegation. Users and the regulatory authority should also be attuned and attentive to these elements against user manipulation by ruthless developers.

9.2 Limitations

While the previous chapters have already detailed the limitations of each study individually, this section assumes a macroscopic perspective and discusses several other limitations that exist consistently throughout our research.

To investigate delegation in critical scenarios, we either gave participants an hypothetical critical context or placed them in a simple setting that emulates real-world cases. These approaches of contextualization may hardly replicate the subjective feelings (e.g., tension or pressure) entailed in critical scenarios. The inadequate simulation of the affective aspect might have inclined participants toward the rational side and consequently decreased the impact of the affective and technological dimensions. Carrying out more field studies would be highly informative to confirm our findings obtained in laboratory or remote settings. Alternatively, one may also consider fully utilizing immersive media technologies to envelope participants in a virtual environment that replicates critical scenarios to fine details. Participants may respond realistically to immersive virtual environments of high-level realism [Slater, 2009]. In several studies (cf. Chapters 4, 6, 7, and 8), we employed a performance-dependent rewarding mechanism to induce the sense of criticality. However, the amount of the reward was too small to be critically consequential, due to which this dynamic rewarding mechanism might not be as effective as expected and, in the worst case, might have gone awry and made the games more fun and thrilling instead of critical. A potentially viable solution is to remove the mechanism and, as mentioned above, immerse participants in a fully virtual environment replicating a critical context, for example, the *trolley problem*.

The visual representation of the virtual agents used in our research was inconsistent. In some cases (cf. Chapter 8 and Experiment 1 in Chapter 7), the agents were embodied in virtual humans, whereas for others (cf. Chapter 4, Chapter 6, and Experiment 2 in Chapter 8) the agents had a robotic representation. The selection of visual appearance was based on our early study presented in Chapter 5, where humanlike characters, including both humans and robots, were found to be the most favorite representation for virtual agents that carry out critical tasks for users. However, the choice of human or robot was mostly arbitrary given their similar user preference. The inconsistency decreased the comparability among different studies, which might have provided more insights into the issue in question.

The interaction between participants and the virtual agents was short and constrained, as exemplified by the ToM experiment (cf. Chapter 6), where the interaction lasted only a few minutes, and there was no communication except for the offer exchange. Due to the limited contact, participants might only have meager information and consequently find it difficult to make their delegation decisions, not to mention the assessment of other factors that develop over time, like trust. This issue lowers the reliability and generalizability of our findings in the context of repeated or long-term interaction. For future studies, one may consider engaging participants in the interaction for an extended period of time.

Delegation to virtual agents was mostly operationalized as a single binary choice in our studies. As previously discussed in Section 7.2.3, this approach is coarse-grained and may overlook factors with a small effect. More fine-grained measures, such as self-reported intention to delegate (cf. Chapter 6) or consecutive delegation decisions (cf. Chapter 7), should be employed. To capture small-effect factors and prevent them from being overshadowed, one may also consider forcing participants to delegate and offering them several agents that differ in the investigated factor.

9.3 Future Work

There are many possible future research avenues. Apart from those mentioned above for addressing the limitations, the list below gives several other paths that may lead to interesting or significant findings.

- **Focusing on less critical scenarios.** Our research indicates that the analytical dimension is highly influential in the delegation of important tasks to virtual agents. However, its influence may degrade in less critical scenarios, allowing the other two dimensions to unfold their impacts on delegation. Thus, studies using less critical contexts have the potential to reveal the impact of the affective and technological dimensions and further clarify the relationships among the three dimensions. Findings derived from these studies also carry practical implication since virtual agents are still mainly used in less critical scenarios such as entertainment and education.

- **Employing multi-agent and multi-task settings.** To isolate and examine the effect of individual factors, our research was mostly conducted in a minimal setting of a dyadic interaction involving a single binary decision. Although this minimal approach is a common practice for empirical studies, users' delegatory behavior may differ in more complex settings. Thus, one may consider using multi-agent and multi-task scenarios to confirm the findings obtained from minimal settings. The results can also provide insights from more sociological and economic perspectives in addition to the psychological stance that this thesis assumes.
- **Devising new instruments.** An obstacle hindering research on user-agent delegation is the lack of a validated approach for measuring the intention to delegate. Devising a questionnaire as such is significantly meaningful since it provides a standard for measuring and comparing the effects of different factors. Similarly, according to our best knowledge, there is no game (as in, e.g., the Prisoners' Dilemma) dedicated to investigating user-agent delegation. Developing such a game can provide a validated common framework for future studies on delegation to virtual agents.

List of Figures

List of Tables

List of Publications

Sun, N., and Botev, J. (2020) Intelligent Adaptive Agents and Trust in Virtual and Augmented Reality. In *Proceedings of the 19th IEEE International Symposium on Mixed and Augmented Reality (ISMAR) Adjunct*, pages 303–305.

Sun, N. and Botev, J. (2021a). Intelligent Autonomous Agents and Trust in Virtual Reality. *Computers in Human Behavior Reports*, 4:100146.

Sun, N. and Botev, J. (2021b). Virtual Agent Representation for Critical Transactions. In *Proceedings of the 13th International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE)*, pages 25–29.

Sun, N. and Botev, J. (2021c). Why Do We Delegate to Intelligent Virtual Agents? Influencing Factors on Delegation Decisions. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI)*, pages 386–390.

Sun, N., Botev, J., Khaluf, Y., and Simoens, P. (2022). Theory of Mind and Delegation to Robotic Virtual Agents. In *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 454–460.

Sun, N. and Botev, J. (2023). Technological Immersion and Delegation to Virtual Agents. *Multimodal Technologies and Interaction*, 7(11):106.

Sun, N., Botev, J., and Simoens, P. (2023). The Effect of Rapport on Delegation to Virtual Agents. In *Proceedings of the 23rd International Conference on Intelligent Virtual Agents (IVA)*.

Bibliography

Abdullah, A. S., Gaehde, S., and Bickmore, T. (2018). A Tablet Based Embodied Conversational Agent to Promote Smoking Cessation Among Veterans: A Feasibility Study. *Journal of Epidemiology And Global Health*, 8(3-4):225–230.

Aburumman, N., Gillies, M., Ward, J. A., and Hamilton, A. F. d. C. (2022). Nonverbal Communication in Virtual Reality: Nodding as a Social Signal in Virtual Interactions. *International Journal of Human-Computer Studies*, 164.

Adams, D., Bah, A., Barwulor, C., Musaby, N., Pitkin, K., and Redmiles, E. M. (2018). Ethics Emerging: The Story of Privacy and Security Perceptions in Virtual Reality. In *Proceedings of the 14th Symposium on Usable Privacy and Security (SOUPS)*, pages 427–442.

Aggarwal, P. and Mazumdar, T. (2008). Decision Delegation: A Conceptualization and Empirical Investigation. *Psychology & Marketing*, 25(1):71–93.

Akinola, M., Martin, A. E., and Phillips, K. W. (2018). To Delegate or Not to Delegate: Gender Differences in Affective Associations and Behavioral Responses to Delegation. *Academy of Management Journal*, 61(4):1467–1491.

Allen, M. S., Iliescu, D., and Greiff, S. (2022). Single Item Measures in Psychological Science. *European Journal of Psychological Assessment*, 38.

Alonso, R. and Matouschek, N. (2007). Relational Delegation. *The RAND Journal of Economics*, 38(4):1070–1089.

Anderson, E. (1985). The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis. *Marketing Science*, 4(3):234–254.

Andrist, S., Pejsa, T., Mutlu, B., and Gleicher, M. (2012). Designing Effective Gaze Mechanisms for Virtual Agents. In *Proceedings of the 30th CHI Conference on Human Factors in Computing Systems*, pages 705–714.

Ang, L. and Eisend, M. (2018). Single Versus Multiple Measurement of Attitudes: A Meta-Analysis of Advertising Studies Validates the Single-Item Measure Approach. *Journal of Advertising Research*, 58(2):218–227.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology*, 63(4):596–612.

Bailenson, J. N., Aharoni, E., Beall, A. C., Guadagno, R. E., Dimov, A., and Blascovich, J. (2004). Comparing Behavioral and Self-Report Measures of Embodied Agents' Social Presence in Immersive Virtual Environments. In *Proceedings of the 7th Annual International Workshop on Presence*.

Baird, A. and Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*, 45(1):315–341.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the Autistic Child Have a "Theory of Mind"? *Cognition*, 21(1):37–46.

Baylor, A. L. (2009). Promoting Motivation With Virtual Agents and Avatars: Role of Visual Presence and Appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3559–3565.

Baylor, A. L. and Kim, S. (2009). Designing Nonverbal Communication for Pedagogical Agents: When Less Is More. *Computers in Human Behavior*, 25(2):450–457.

Bee, N., André, E., and Tober, S. (2009). Breaking the Ice in Human-Agent Communication: Eye-Gaze Based Initiation of Contact With an Embodied

Conversational Agent. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA)*, pages 229–242.

Bergmann, K., Eyssel, F., and Kopp, S. (2012). A Second Chance to Make a First Impression? How Appearance and Nonverbal Behavior Affect Perceived Warmth and Competence of Virtual Agents Over Time. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents (IVA)*, pages 126–138.

Bergmann, K., Kopp, S., and Eyssel, F. (2010). Individualized Gesturing Outperforms Average Gesturing – Evaluating Gesture Production in Virtual Humans. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA)*, pages 104–117.

Berry, D. S., Pennebaker, J. W., Mueller, J. S., and Hiller, W. S. (1997). Linguistic Bases of Social Perception. *Personality and Social Psychology Bulletin*, 23(5):526–537.

Bigman, Y. E. and Gray, K. (2018). People Are Averse to Machines Making Moral Decisions. *Cognition*, 181:21–34.

Blum-Kulka, S. and Dvir-Gvirsman, S. (2010). Peer Interaction and Learning. In Peterson, P., Baker, E., and McGaw, B., editors, *International Encyclopedia of Education*, pages 444–449. Elsevier, third edition.

Bobadilla-Suarez, S., Sunstein, C. R., and Sharot, T. (2017). The Intrinsic Value of Choice: The Propensity to Under-Delegate in the Face of Potential Gains and Losses. *Journal of Risk and Uncertainty*, 54(3):187–202.

Bombari, D., Schmid M., M., Canadas, E., and Bachmann, M. (2015). Studying Social Interactions Through Immersive Virtual Environment Technology: Virtues, Pitfalls, and Future Challenges. *Frontiers in Psychology*, 6.

Boucaud, F., Tafiani, Q., Pelachaud, C., and Thouvenin, I. (2019). Social Touch in Human-Agent Interactions in an Immersive Virtual Environment. In *Proceedings of the 3rd International Conference on Human Computer Interaction Theory and Applications (HUCAPP)*, pages 129–136.

Bronstein, I., Nelson, N., Livnat, Z., and Ben-Ari, R. (2012). Rapport in Negotiation: The Contribution of the Verbal Channel. *Journal of Conflict Resolution*, 56(6):1089–1115.

Buchholz, V., Kulms, P., and Kopp, S. (2017). It's (Not) Your Fault! Blame and Trust Repair in Human-Agent Cooperation. In *Proceedings of the 6th Interdisziplinäre Workshop-Reihe "Kognitive Systeme"*.

Buisine, S., Abrilian, S., and Martin, J.-C. (2004). Evaluation of Multimodal Behaviour of Embodied Agents: Cooperation Between Speech and Gestures. In Ruttkay, Z. and Pelachaud, C., editors, *From Brows to Trust: Evaluating Embodied Conversational Agents*, pages 217–238. Springer Netherlands.

Buttussi, F. and Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1063–1076.

Cai, S., Ke, P., Narumi, T., and Zhu, K. (2020). Thermairglove: A Pneumatic Glove for Thermal Perception and Material Identification in Virtual Reality. In *Proceedings of the 27th IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)*, pages 248–257.

Candrian, C. and Scherer, A. (2022). Rise of the Machines: Delegating Decisions to Autonomous AI. *Computers in Human Behavior*, 134:107308.

Cao, S., Nandakumar, K., Babu, R., and Thompson, B. (2020). Game Play in Virtual Reality Driving Simulation Involving Head-Mounted Display and Comparison to Desktop Display. *Virtual Reality*, 24(3):503–513.

Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., and Göbel, S. (2021). Cybersickness in Current-Generation Virtual Reality Head-Mounted Displays: Systematic Review and Outlook. *Virtual Reality*, 25(4):1153–1170.

Cassell, J. (2004). Towards a Model of Technology and Literacy Development: Story Listening Systems. *Journal of Applied Developmental Psychology*, 25(1):75–105.

Cassell, J. (2022). Socially Interactive Agents as Peers. In Lugrin, B., Pelachaud, C., and Traum, D., editors, *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, pages 331–366. Association for Computing Machinery, first edition.

Castelfranchi, C. and Falcone, R. (1998). Towards a Theory of Delegation for Agent-Based Systems. *Robotics and Autonomous Systems*, 24(3-4):141–157.

Cerekovic, A., Aran, O., and Gatica-Perez, D. (2014). How Do You Like Your Virtual Agent?: Human-Agent Interaction Experience Through Nonverbal Features and Personality Traits. In *Proceedings of the 5th International Workshop on Human Behavior Understanding (HBU)*, pages 1–15.

Chiou, E. K., Schroeder, N. L., and Craig, S. D. (2020). How We Trust, Perceive, and Learn From Virtual Humans: The Influence of Voice Quality. *Computers & Education*, 146:103756.

Cho, M.-G. (2021). A Study on Augmented Reality-Based Virtual Pets for the Elderly Living Alone. In *Proceedings of the 12th International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1280–1283.

Chowdhury, T. I., Ferdous, S. M. S., and Quarles, J. (2021). VR Disability Simulation Reduces Implicit Bias Towards Persons With Disabilities. *IEEE Transactions on Visualization and Computer Graphics*, 27(6):3079–3090.

Clark, R. E. and Choi, S. (2005). Five Design Principles for Experiments on the Effects of Animated Pedagogical Agents. *Journal of Educational Computing Research*, 32(3):209–225.

Cook, K. S., Yamagishi, T., Cheshire, C., Cooper, R., Matsuda, M., and Mashima, R. (2005). Trust Building via Risk Taking: A Cross-Societal Experiment. *Social Psychology Quarterly*, 68(2):121–142.

Şerban, O., Barange, M., Zojaji, S., Pauchet, A., Richard, A., and Chanoni, E. (2017). Interactive Narration With a Child: Impact of Prosody and Facial

Expressions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI)*, pages 23–31.

Cummings, J. J. and Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, 19(2):272–309.

Dai, Z. and MacDorman, K. F. (2018). The Doctor's Digital Double: How Warmth, Competence, and Animation Promote Adherence Intention. *PeerJ Computer Science*, 4:e168.

Daronnat, S., Azzopardi, L., Halvey, M., and Dubiel, M. (2020). Impact of Agent Reliability and Predictability on Trust in Real Time Human-Agent Collaboration. In *Proceedings of the 8th International Conference on Human-Agent Interaction (HAI)*, pages 131–139.

Davis, R. and Antonenko, P. (2017). Effects of Pedagogical Agent Gestures on Social Acceptance and Learning: Virtual Real Relationships in an Elementary Foreign Language Classroom. *Journal of Interactive Learning Research*, 28(4):459–480.

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., and Parasuraman, R. (2016). Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. *Journal of Experimental Psychology: Applied*, 22(3):331–349.

de Visser, E. J. and Parasuraman, R. (2011). Adaptive Aiding of Human-Robot Teaming: Effects of Imperfect Automation on Performance, Trust, and Workload. *Journal of Cognitive Engineering and Decision Making*, 5(2):209–231.

de Weerd, H., Verbrugge, R., and Verheij, B. (2017). Negotiating With Other Minds: The Role of Recursive Theory of Mind in Negotiation With Incomplete Information. *Autonomous Agents and Multi-Agent Systems*, 31(2):250–287.

DeBruine, L. M. (2002). Facial Resemblance Enhances Trust. *Proceedings of the Royal Society B: Biological Sciences*, 269(1498):1307–1312.

Demiris, G., Thompson, H. J., Lazar, A., and Lin, S.-Y. (2016). Evaluation of a Digital Companion for Older Adults With Mild Cognitive Impairment. In *Proceedings of American Medical Informatics Association Annual Symposium (AMIA)*, volume 2016, pages 496–503.

Deutsch, M. (1962). Cooperation and Trust: Some Theoretical Notes. In Jones, M. R., editor, *Nebraska Symposium on Motivation*, pages 275–320. Universiy of Nebraska Press.

Deutsch, M. (1973). *The Resolution of Conflict: Constructive and Destructive Processes*. Yale University Press.

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., and Marchetti, A. (2020). Shall I Trust You? From Child–Robot Interaction to Trusting Relationships. *Frontiers in Psychology*, 11.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General*, 144(1):114–126.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, 64(3):1155–1170.

Dixon, S. R. and Wickens, C. D. (2006). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors*, 48(3):474–486.

DiYanni, C., Nini, D., Rheel, W., and Livelli, A. (2012). 'I Won't Trust You If I Think You're Trying to Deceive Me': Relations Between Selective Trust, Theory of Mind, and Imitation in Early Childhood. *Journal of Cognition and Development*, 13(3):354–371.

Doirado, E. and Martinho, C. (2010). I Mean It! Detecting User Intentions to Create Believable Behaviour for Virtual Agents in Games. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 83–90.

Dowsett, A. and Jackson, M. (2019). The Effect of Violence and Competition Within Video Games on Aggression. *Computers in Human Behavior*, 99:22–27.

Dunn, J. R. and Schweitzer, M. E. (2005). Feeling and Believing: The Influence of Emotion on Trust. *Journal of Personality and Social Psychology*, 88(5):736–748.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. (2003). The Role of Trust in Automation Reliance. *International Journal of Human-Computer Studies*, 58(6):697–718.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., and Dawe, L. A. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors*, 44(1):79–94.

Eisenhardt, K. M. (1985). Control: Organizational and Economic Approaches. *Management Science*, 31(2):134–149.

Eisenhardt, K. M. (1988). Agency- and Institutional-Theory Explanations: The Case of Retail Sales Compensation. *The Academy of Management Journal*, 31(3):488–511.

Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *Academy of Management Review*, 14(1):57–74.

Ermi, L. and Mäyrä, F. (2005). Fundamental Components of the Gameplay Experience: Analyzing Immersion. In *Proceedings of the 2nd DiGRA International Conference*, pages 15–57.

Faita, C., Vanni, F., Tanca, C., Ruffaldi, E., Carrozzino, M., and Bergamasco, M. (2016). Investigating the Process of Emotion Recognition in Immersive and Non-Immersive Virtual Technological Setups. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (VRST)*, pages 61–64.

Fama, E. F. (1980). Agency Problems and the Theory of the Firm. *Journal of Political Economy*, 88(2):288–307.

Fama, E. F. and Jensen, M. C. (1983). Separation of Ownership and Control. *The Journal of Law and Economics*, 26(2):301–325.

Feng, Y., Duives, D. C., and Hoogendoorn, S. P. (2022). Wayfinding Behaviour in a Multi-Level Building: A Comparative Study of HMD VR and Desktop VR. *Advanced Engineering Informatics*, 51.

Fernández Domingos, E., Terrucha, I., Suchon, R., Grujić, J., Burguillo, J. C., Santos, F. C., and Lenaerts, T. (2022). Delegation to Artificial Agents Fosters Prosocial Behaviors in the Collective Risk Dilemma. *Scientific Reports*, 12(1).

Fogg, B. J. (2009). A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology (PERSUASIVE)*, pages 1–7.

Friedman, B., Khan Jr., P. H., and Howe, D. C. (2000). Trust Online. *Communications of the ACM*, 43(12):34–40.

Fügener, A., Grahl, J., Gupta, A., and Ketter, W. (2022). Cognitive Challenges in Human-Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*, 33(2):678–696.

Fügener, A., Grahl, J., Gupta, A., Ketter, W., and Taudien, A. (2021). Exploring User Heterogeneity in Human Delegation Behavior Towards AI. In *Proceedings of the 42nd International Conference on Information Systems (ICIS)*.

Fukuyama, F. (1996). *Trust: The Social Virtues and the Creation of Prosperity*. Simon and Schuster.

Gal, Y., Grosz, B. J., Kraus, S., Pfeffer, A., and Shieber, S. (2005). Colored Trails: A Formalism for Investigating Decision-Making in Strategic Environments. In *Proceedings of the 2005 IJCAI Workshop on Reasoning, Representation, and Learning in Computer Games*, pages 25–30.

Gallace, A. and Girondini, M. (2022). Social Touch in Virtual Reality. *Current Opinion in Behavioral Sciences*, 43:249–254.

Gamito, P., Oliveira, J., Santos, N., Pacheco, J., Morais, D., Saraiva, T., Soares, F., Mayor, C. S., and Barata, A. F. (2014). Virtual Exercises to Promote Cognitive Recovery in Stroke Patients: The Comparison Between Head Mounted Displays Versus Screen Exposure Methods. *International Journal on Disability and Human Development*, 13(3):337–342.

Gan, Q., Liu, Z., Liu, T., Zhao, Y., and Chai, Y. (2022). Design and User Experience Analysis of AR Intelligent Virtual Agents in Smartphone. *Cognitive Systems Research*, 78:33–47.

Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). The Impact of Eye Gaze on Communication Using Humanoid Avatars. In *Proceedings of the 19th CHI Conference on Human Factors in Computing Systems*, pages 309–316.

Garcia, Y., Khooshabeh, P., and Ouimette, B. (2016). The Effect of a Virtual Agent's Emotional Facial Expressions on the Mind's Eye Test. In *Proceedings of the 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 83–87.

Glikson, E. and Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, 14(2):627–660.

Gong, L. and Nass, C. (2007). When a Talking-Face Computer Agent Is Half-Human and Half-Humanoid: Human Identity and Consistency Preference. *Human Communication Research*, 33(2):163–193.

Gratch, J. and Lucas, G. (2021). Rapport Between Humans and Socially Interactive Agents. In Lugrin, B., Pelachaud, C., and Traum, D., editors, *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume*

1: Methods, Behavior, Cognition, pages 433–462. Association for Computing Machinery, first edition.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., and Morency, L.-P. (2006). Virtual Rapport. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA)*, pages 14–27.

Grossman, S. J. and Hart, O. D. (1992). An Analysis of the Principal-Agent Problem. In Dionne, G. and Harrington, S. E., editors, *Foundations of Insurance Economics: Readings in Economics and Finance*, pages 302–340. Springer Netherlands.

Grosz, B., Kraus, S., Talman, S., Stossel, B., and Havlin, M. (2004). The Influence of Social Dependencies on Decision-Making: Initial Investigations With a New Game. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 782–789.

Guggisberg, M. (2020). Sexually Explicit Video Games and Online Pornography – The Promotion of Sexual Violence: A Critical Commentary. *Aggression and Violent Behavior*, 53:101432.

Guimarães, M., Prada, R., Santos, P. A., Dias, J., Jhala, A., and Mascarenhas, S. (2020). The Impact of Virtual Reality in the Social Presence of a Virtual Agent. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA)*.

Gulz, A. and Haake, M. (2006). Visual Design of Virtual Pedagogical Agents: Naturalism Versus Stylization in Static Appearance. In *Proceedings of the 3rd International Design and Engagability Conference (iDec)*.

Gur, N. and Bjørnskov, C. (2017). Trust and Delegation: Theory and Evidence. *Journal of Comparative Economics*, 45(3):644–657.

Hald, K., Weitz, K., André, E., and Rehm, M. (2021). “An Error Occurred!” – Trust Repair With Virtual Robot Using Levels of Mistake Explanation. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI)*, pages 218–226.

Hanna, N. and Richards, D. (2018). The Impact of Multimodal Communication on a Shared Mental Model, Trust, and Commitment in Human-Intelligent Virtual Agent Teams. *Multimodal Technologies and Interaction*, 2(3).

Harbers, M., Peeters, M. M. M., and Neerincx, M. A. (2017). Perceived Autonomy of Robots: Effects of Appearance and Context. In *Proceedings of the 2015 International Conference on Robot Ethics (ICRE)*, pages 19–33.

Harper, R. G., Wiens, A. N., and Matarazzo, J. D. (1978). *Nonverbal Communication: The State of the Art*. John Wiley & Sons.

Harrigan, J. A., Oxman, T. E., and Rosenthal, R. (1985). Rapport Expressed Through Nonverbal Behavior. *Journal of Nonverbal Behavior*, 9:95–110.

Hattab, G., Hatzipanayioti, A., Klimova, A., Pfeiffer, M., Klausing, P., Breucha, M., Bechtolsheim, F. v., Helmert, J. R., Weitz, J., Pannasch, S., and Speidel, S. (2021). Investigating the Utility of VR for Spatial Understanding in Surgical Planning: Evaluation of Head-Mounted to Desktop Display. *Scientific Reports*, 11(1).

Henderson, S. J. and Feiner, S. (2009). Evaluating the Benefits of Augmented Reality for Task Localization in Maintenance of an Armored Personnel Carrier Turret. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 135–144.

Hendrick, C. (1972). Effects of Salience of Stimulus Inconsistency on Impression Formation. *Journal of Personality and Social Psychology*, 22(2):219–222.

Hepperle, D., Ödell, H., and Wölfel, M. (2020). Differences in the Uncanny Valley Between Head-Mounted Displays and Monitors. In *Proceedings of the 19th International Conference on Cyberworlds (CW)*, pages 41–48.

Hepperle, D., Purps, C. F., Deuchler, J., and Wölfel, M. (2022). Aspects of Visual Avatar Appearance: Self-Representation, Display Type, and Uncanny Valley. *The Visual Computer*, 38:1227–1244.

Heylen, D., van Es, I., Nijholt, A., and van Dijk, B. (2002). Experimenting With the Gaze of a Conversational Agent. In *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, pages 93–100.

Hoff, K. A. and Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors*, 57(3):407–434.

Hoogendoorn, M. and Soumokil, J. (2010). Evaluation of Virtual Agents Utilizing Theory of Mind in a Real Time Action Game. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 59–66.

Hoppe, M., Rossmy, B., Neumann, D. P., Streuber, S., Schmidt, A., and Machulla, T.-K. (2020). A Human Touch: Social Touch Increases the Perceived Human-Likeness of Agents in Virtual Reality. In *Proceedings of the 38th CHI Conference on Human Factors in Computing Systems*, pages 1–11.

Hoque, M., Courgeon, M., Martin, J.-C., Mutlu, B., and Picard, R. W. (2013). Mach: My Automated Conversation Coach. In *Proceedings of the 15th ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 697–706.

Hosseinpanah, A., Krämer, N. C., and Straßmann, C. (2018). Empathy for Everyone? The Effect of Age When Evaluating a Virtual Agent. In *Proceedings of the 6th International Conference on Human-Agent Interaction (HAI)*, pages 184–190.

Hove, M. J. and Risen, J. L. (2009). It's All in the Timing: Interpersonal Synchrony Increases Affiliation. *Social Cognition*, 27(6):949–961.

Huisman, G., Kolkmeier, J., and Heylen, D. (2014). Simulated Social Touch in a Collaborative Game. In *Proceedings of the 9th International Conference on Human Haptic Sensing and Touch Enabled Computer Applications (EuroHaptics)*, pages 248–256.

Hukal, P., Berente, N., Germonprez, M., and Schecter, A. (2019). Bots Coordinating Work in Open Source Software Projects. *Computer*, 52(9):52–60.

Hvass, J., Larsen, O., Vendelbo, K., Nilsson, N., Nordahl, R., and Serafin, S. (2017). Visual Realism and Presence in a Virtual Reality Game. In *Proceedings of the 11th 3DTV Conference: The True Vision – Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4.

Jenks, J. M. and Kelly, J. M. (1985). *Don't Do, Delegate!* Franklin Watts.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and Defining the Experience of Immersion in Games. *International Journal of Human-Computer Studies*, 66(9):641–661.

Jensen, M. C. (1986). Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *The American Economic Review*, 76(2):323–329.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3(4):305–360.

Jian, J.-Y., Bisantz, A. M., and Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4(1):53–71.

Johnson, J. D., Sanchez, J., Fisk, A. D., and Rogers, W. A. (2004). Type of Automation Failure: The Effects on Trust and Reliance in Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(18):2163–2167.

Johnson, W. L. and Rickel, J. (1997). Steve: An Animated Pedagogical Agent for Procedural Training in Virtual Environments. *ACM SIGART Bulletin*, 8(1-4):16–21.

Karacora, B., Dehghani, M., Krämer-Mertens, N., and Gratch, J. (2012). The Influence of Virtual Agents' Gender and Rapport on Enhancing Math Performance. pages 563–568.

Keef, C. V., Kayser, L. V., Tronboll, S., Carpenter, C. W., Root, N. B., Finn III, M., O'Connor, T. F., Abuhamdieh, S. N., Davies, D. M., Runser, R., et al. (2020). Virtual Texture Generated Using Elastomeric Conductive Block Copolymer in a Wireless Multimodal Haptic Glove. *Advanced Intelligent Systems*, 2(4).

Khalid, H. M., Shiung, L. W., Sheng, V. B., and Helander, M. G. (2018). Trust of Virtual Agent in Multi Actor Interactions. *Journal of Robotics, Networking and Artificial Life*, 4(4):295–298.

Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., and Welch, G. F. (2018). Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR. In *Proceedings of the 17th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 105–114.

Kim, K., Rosenthal, M. Z., Zielinski, D. J., and Brady, R. (2014). Effects of Virtual Environment Platforms on Emotional Responses. *Computer Methods and Programs in Biomedicine*, 113(3):882–893.

Klein, A., Yumak, Z., Beij, A., and van der Stappen, A. F. (2019). Data-Driven Gaze Animation Using Recurrent Neural Networks. In *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pages 1–11.

Klingenberg, S., Jørgensen, M. L. M., Dandanell, G., Skriver, K., Mottelson, A., and Makransky, G. (2020). Investigating the Effect of Teaching as a Generative Learning Strategy When Learning Through Desktop and Immersive VR: A Media and Methods Experiment. *British Journal of Educational Technology*, 51(6):2115–2138.

Koda, T. and Maes, P. (1996). Agents With Faces: The Effect of Personification. In *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication (RO-MAN)*, pages 189–194.

Krämer, N. C., Karacora, B., Lucas, G., Dehghani, M., Rüther, G., and Gratch, J. (2016). Closing the Gender Gap in STEM With Friendly Male Instructors?

On the Effects of Rapport Behavior and Gender of a Virtual Agent in an Instructional Interaction. *Computers & Education*, 99:1–13.

Krämer, N. C., Lucas, G., Schmitt, L., and Gratch, J. (2018). Social Snacking With a Virtual Agent – On the Interrelation of Need to Belong and Effects of Social Responsiveness when Interacting with Artificial Entities. *International Journal of Human-Computer Studies*, 109:112–121.

Krämer, N. C., Simons, N., and Kopp, S. (2007). The Effects of an Embodied Conversational Agent's Nonverbal Behavior on User's Evaluation and Behavioral Mimicry. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA)*, pages 238–251.

Krokos, E., Plaisant, C., and Varshney, A. (2019). Virtual Memory Palaces: Immersion Aids Recall. *Virtual Reality*, 23(1):1–15.

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., and Kappas, A. (2007). Facial Dynamics as Indicators of Trustworthiness and Cooperative Behavior. *Emotion*, 7(4):730–735.

Lane, H. C., Cahill, C., Foutz, S., Auerbach, D., Noren, D., Lussenhop, C., and Swartout, W. (2013). The Effects of a Pedagogical Agent for Informal Science Education on Learner Behaviors and Self-Efficacy. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED)*, pages 309–318.

Lane, H. C. and Schroeder, N. L. (2022). Pedagogical Agents. In Lugrin, B., Pelachaud, C., and Traum, D., editors, *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, pages 307–330. Association for Computing Machinery, first edition.

Latu, I. M., Mast, M. S., Lammers, J., and Bombari, D. (2013). Successful Female Leaders Empower Women's Behavior in Leadership Tasks. *Journal of Experimental Social Psychology*, 49(3):444–448.

Launay, J., Dean, R. T., and Bailes, F. (2013). Synchronization Can Influence Trust Following Virtual Interaction. *Experimental Psychology*, 60(1):53–63.

LaViola Jr., J. J., Kruijff, E., McMahan, R. P., Bowman, D. A., and Poupyrev, I. (2017). *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, second edition.

Leana, C. R. (1987). Power Relinquishment Versus Power Sharing: Theoretical Clarification and Empirical Comparison of Delegation and Participation. *Journal of Applied Psychology*, 72(2):228–233.

Lecce, S., Ceccato, I., Bianco, F., Rosi, A., Bottiroli, S., and Cavallini, E. (2017). Theory of Mind and Social Relationships in Older Adults: The Role of Social Motivation. *Aging & Mental Health*, 21(3):253–258.

Lee, J. and Moray, N. (1992). Trust, Control Strategies and Allocation of Function in Human-Machine Systems. *Ergonomics*, 35(10):1243–1270.

Lee, J. D. and See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1):50–80.

Lee, J.-E. R., Nass, C., Brave, S. B., Morishima, Y., Nakajima, H., and Yamada, R. (2007). The Case for Caring Colearners: The Effects of a Computer-Mediated Colearner Agent on Trust and Learning. *Journal of Communication*, 57(2):183–204.

Lehdonvirta, V. and Castronova, E. (2014). *Virtual Economies: Design and Analysis*. MIT Press.

Lewis, J. D. and Weigert, A. (1985). Trust as a Social Reality. *Social Forces*, 63(4):967–985.

Leyer, M., Aysolmaz, B., and Iren, D. (2021). Acceptance of AI for Delegating Emotional Intelligence: Results From an Experiment. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*, pages 6307–6316.

Li, C., Androulakaki, T., Gao, A. Y., Yang, F., Saikia, H., Peters, C., and Skantze, G. (2018). Effects of Posture and Embodiment on Social Distance in Human-Agent Interaction in Mixed Reality. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA)*, pages 191–196.

Li, J. (2015). The Benefit of Being Physically Present: A Survey of Experimental Works Comparing Copresent Robots, Telepresent Robots and Virtual Agents. *International Journal of Human-Computer Studies*, 77:23–37.

Liehner, G. L., Brauner, P., Schaar, A. K., and Ziefle, M. (2021). Delegation of Moral Tasks to Automated Agents – The Impact of Risk and Context on Trusting a Machine to Perform a Task. *IEEE Transactions on Technology and Society*, 3(1):46–57.

Lim, J., Mountstephens, J., and Teo, J. (2020). Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges. *Sensors*, 20(8).

Löffler, D., Dörrenbächer, J., and Hassenzahl, M. (2020). The Uncanny Valley Effect in Zoomorphic Robots: The U-Shaped Relation Between Animal Likeness and Likeability. In *Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 261–270.

Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103.

Lubars, B. and Tan, C. (2019). Ask Not What AI Can Do, But What AI Should Do: Towards a Framework of Task Delegability. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*, pages 57–67.

Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. (2014). It's Only a Computer: Virtual Humans Increase Willingness to Disclose. *Computers in Human Behavior*, 37:94–100.

Lugrin, B. (2021). Introduction to Socially Interactive Agents. In Lugrin, B., Pelachaud, C., and Traum, D., editors, *The Handbook on Socially Interactive*

Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition, pages 1–20. Association for Computing Machinery, first edition.

Lugrin, B., Pelachaud, C., and Traum, D., editors (2021). *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*. Association for Computing Machinery, first edition.

Lugrin, B., Pelachaud, C., and Traum, D., editors (2022). *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*. Association for Computing Machinery, first edition.

Luhmann, N. (1979). *Trust and Power*. John Wiley & Sons.

Luo, L., Weng, D., Ding, N., Hao, J., and Tu, Z. (2022). The Effect of Avatar Facial Expressions on Trust Building in Social Virtual Reality. *The Visual Computer*, pages 1–14.

Lupia, A. (2001). Delegation of Power: Agency Theory. In *International Encyclopedia of the Social and Behavioral Sciences*, pages 3375–3377. Pergamon.

Madeira, T., Marques, B., Neves, P., Dias, P., and Santos, B. S. (2022). Comparing Desktop vs. Mobile Interaction for the Creation of Pervasive Augmented Reality Experiences. *Journal of Imaging*, 8(3).

Madhavan, P., Wiegmann, D. A., and Lacson, F. C. (2006). Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids. *Human Factors*, 48(2):241–256.

Makransky, G., Terkildsen, T. S., and Mayer, R. E. (2019). Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence But Less Learning. *Learning and Instruction*, 60:225–236.

Manzey, D., Reichenbach, J., and Onnasch, L. (2012). Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making*, 6(1):57–87.

Marsh, S. P. (1994). *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling.

Martínez-Miranda, J., Martínez, A., Ramos, R., Aguilar, H., Jiménez, L., Arias, H., Rosales, G., and Valencia, E. (2019). Assessment of Users' Acceptability of a Mobile-Based Embodied Conversational Agent for the Prevention and Detection of Suicidal Behaviour. *Journal of Medical Systems*, 43(8).

Matsui, T. and Koike, A. (2021). Who Is to Blame? The Appearance of Virtual Agents and the Attribution of Perceived Responsibility. *Sensors*, 21(8).

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3):709–734.

McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render Me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. *ACM Transactions on Graphics*, 31(4):1–11.

McMahan, A. (2013). Immersion, Engagement, and Presence: A Method for Analyzing 3-D Video Games. In Wolf, M. J. and Perron, B., editors, *The Video Game Theory Reader*, pages 67–86. Routledge.

Mehrabian, A. and Ferris, S. R. (1967). Inference of Attitudes From Nonverbal Communication in Two Channels. *Journal of Consulting Psychology*, 31(3):248–252.

Mehrabian, A. and Wiener, M. (1967). Decoding of Inconsistent Communications. *Journal of Personality and Social Psychology*, 6(1):109–114.

Melson, G. F., Kahn, P. H., Beck, A. M., Friedman, B., Roberts, T., and Garrett, E. (2005). Robots as Dogs? Children's Interactions With the Robotic Dog AIBO and a Live Australian Shepherd. In *Proceedings of the 23rd CHI Conference on Human Factors in Computing Systems*, pages 1649–1652.

Miles, L. K., Nind, L. K., and Macrae, C. N. (2009). The Rhythm of Rapport: Interpersonal Synchrony and Social Perception. *Journal of Experimental Social Psychology*, 45(3):585–589.

Milewski, A. E. and Lewis, S. H. (1997). Delegating to Software Agents. *International Journal of Human-Computer Studies*, 46(4):485–500.

Miller, H. L. and Bugnariu, N. L. (2016). Level of Immersion in Virtual Environments Impacts the Ability to Assess and Teach Social Skills in Autism Spectrum Disorder. *Cyberpsychology, Behavior, and Social Networking*, 19(4):246–256.

Monroe, S. and Vangsness, L. (2022). The Effects of Task Difficulty and Stress on Trust in an Automated Navigation Aid. In *Proceedings of the 66th Human Factors and Ergonomics Society Annual Meeting (HFES)*, pages 1080–1084.

Montoya, R. M., Horton, R. S., and Kirchner, J. (2008). Is Actual Similarity Necessary for Attraction? A Meta-Analysis of Actual and Perceived Similarity. *Journal of Social and Personal Relationships*, 25(6):889–922.

Moradinezhad, R. and Solovey, E. T. (2021). Investigating Trust in Interaction With Inconsistent Embodied Virtual Agents. *International Journal of Social Robotics*, 13(8):2103–2118.

Moreno, R. and Mayer, R. E. (2004). Personalized Messages That Promote Science Learning in Virtual Environments. *Journal of Educational Psychology*, 96(1):165–173.

Mori, M. (1970). Bukimi no tani [The Uncanny Valley]. *Energy*, 7:33–35.

Morris, B., Doshi, A., and Trivedi, M. (2011). Lane Change Intent Prediction for Driver Assistance: On-Road Design and Evaluation. In *Proceedings of the 16th IEEE Intelligent Vehicles Symposium (IEEE IV)*, pages 895–901.

Morrow Jr., J. L., Hansen, M. H., and Pearson, A. W. (2004). The Cognitive and Affective Antecedents of General Trust Within Cooperative Organizations. *Journal of Managerial Issues*, 16:48–64.

Mou, W., Ruocco, M., Zanatto, D., and Cangelosi, A. (2020). When Would You Trust a Robot? A Study on Trust and Theory of Mind in Human-Robot Interactions. In *Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 956–962.

Muir, B. M. (1994). Trust in Automation: Part I. Theoretical Issues in the Study of Trust and Human Intervention in Automated Systems. *Ergonomics*, 37(11):1905–1922.

Murray, J. H. (2017). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. MIT Press, second edition.

Murray, M., Walker, N., Nanavati, A., Alves-Oliveira, P., Filippov, N., Sauppe, A., Mutlu, B., and Cakmak, M. (2022). Learning Backchanneling Behaviors for a Social Robot via Data Augmentation From Human-Human Conversations. In *Proceedings of the 5th Conference on Robot Learning (CoRL)*, pages 513–525.

Myers, C. D. and Tingley, D. (2016). The Influence of Emotion on Trust. *Political Analysis*, 24(4):492–500.

Narang, S., Best, A., and Manocha, D. (2019). Inferring User Intent Using Bayesian Theory of Mind in Shared Avatar-Agent Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2113–2122.

Nass, C. and Lee, K. M. (2001). Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction. *Journal of Experimental Psychology: Applied*, 7(3):171–181.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers Are Social Actors. In *Proceedings of the 12th CHI Conference on Human Factors in Computing Systems*, pages 72–78.

Natarajan, M. and Gombolay, M. (2020). Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 33–42.

Neff, M., Wang, Y., Abbott, R., and Walker, M. (2010). Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA)*, pages 222–235.

Nilsson, N. C., Nordahl, R., and Serafin, S. (2016). Immersion Revisited: A Review of Existing Definitions of Immersion and Their Relation to Different Theories of Presence. *Human Technology*, 12(2):108–134.

Normoyle, A., Badler, J. B., Fan, T., Badler, N. I., Cassol, V. J., and Musse, S. R. (2013). Evaluating Perceived Trust From Procedurally Animated Gaze. In *Proceedings of the 6th International Conference on Motion in Games (MIG)*, pages 141–148.

Obaid, M., Damian, I., Kistler, F., Endrass, B., Wagner, J., and André, E. (2012). Cultural Behaviors of Virtual Agents in an Augmented Reality Environment. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents (IVA)*, pages 412–418.

Ochs, M., Mestre, D., de Montcheuil, G., Pergandi, J.-M., Saubesty, J., Lombardo, E., Francon, D., and Blache, P. (2019). Training Doctors' Social Skills to Break Bad News: Evaluation of the Impact of Virtual Environment Displays on the Sense of Presence. *Journal on Multimodal User Interfaces*, 13:41–51.

Ochs, M., Niewiadomski, R., and Pelachaud, C. (2010). How a Virtual Agent Should Smile? Morphological and Dynamic Characteristics of Virtual Agent's Smiles. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA)*, pages 427–440.

Odenthal, B., Mayer, M. P., Kabuß, W., and Schlick, C. M. (2014). A Comparative Study of Head-Mounted and Table-Mounted Augmented Vision Systems for Assembly Error Detection. *Human Factors And Ergonomics in Manufacturing & Service Industries*, 24(1):105–123.

Olafsson, S., O'Leary, T. K., and Bickmore, T. W. (2020a). Motivating Health Behavior Change With Humorous Virtual Agents. In *Proceedings of the 20th International Conference on Intelligent Virtual Agents (IVA)*.

Olafsson, S., Wallace, B. C., and Bickmore, T. W. (2020b). Towards a Computational Framework for Automating Substance Use Counseling With Virtual Agents. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 966–974.

Olmos-Raya, E., Ferreira-Cavalcanti, J., Contero, M., Castellanos, M. C., Giglioli, I. A. C., and Alcañiz, M. (2018). Mobile Virtual Reality as an Educational Platform: A Pilot Study on the Impact of Immersion and Positive Emotion Induction in the Learning Process. *Journal of Mathematics, Science and Technology Education*, 14(6):2045–2057.

OpenAI (2023). Introducing ChatGPT. https://openai.com/blog/chatgpt. Last accessed: 2023-05-07.

Pallavicini, F. and Pepe, A. (2019). Comparing Player Experience in Video Games Played in Virtual Reality or on Desktop Displays: Immersion, Flow, and Positive Emotions. In *Proceedings of the 19th Annual Symposium on Computer-Human Interaction in Play (CHI PLAY)*.

Pan, X. and Hamilton, A. F. d. 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.

Parasuraman, R. and Miller, C. A. (2004). Trust and Etiquette in High-Criticality Automated Systems. *Communications of the ACM*, 47(4):51–55.

Parasuraman, R. and Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2):230–253.

Patzer, G. L. (1983). Source Credibility as a Function of Communicator Physical Attractiveness. *Journal of Business Research*, 11(2):229–241.

Payne, J., Szymkowiak, A., Robertson, P., and Johnson, G. (2013). Gendering the Machine: Preferred Virtual Assistant Gender and Realism in Self-Service. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents (IVA)*, pages 106–115.

Pei, G. and Li, T. (2021). A Literature Review of EEG-Based Affective Computing in Marketing. *Frontiers in Psychology*, 12:602843.

Pejsa, T., Gleicher, M., and Mutlu, B. (2017). Who, Me? How Virtual Agents Can Shape Conversational Footing in Virtual Reality. In *Proceedings of the 17th International Conference on Intelligent Virtual Agents (IVA)*, pages 347–359.

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

Perez-Osorio, J., Wiese, E., and Wykowska, A. (2021). Theory of Mind and Joint Attention. In Lugrin, B., Pelachaud, C., and Traum, D., editors, *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, pages 311–348. Association for Computing Machinery, first edition.

Perkis, A., Timmerer, C., et al. (2020). QUALINET White Paper on Definitions of Immersive Media Experience (IMEx). *European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET Meeting (online).* May 25, 2020.

Perlin, K. and Seidman, G. (2008). Autonomous Digital Actors. In *Proceedings of the 1st International Workshop on Motion in Games (MIG)*, pages 246–255.

Peukert, C., Pfeiffer, J., Meißner, M., Pfeiffer, T., and Weinhardt, C. (2019). Shopping in Virtual Reality Stores: The Influence of Immersion on System Adoption. *Journal of Management Information Systems*, 36(3):755–788.

Plechatá, A., Sahula, V., Fayette, D., and Fajnerová, I. (2019). Age-Related Differences With Immersive and Non-immersive Virtual Reality in Memory Assessment. *Frontiers in Psychology*, 10.

Pollard, K. A., Oiknine, A. H., Files, B. T., Sinatra, A. M., Patton, D., Ericson, M., Thomas, J., and Khooshabeh, P. (2020). Level of Immersion Affects Spatial Learning in Virtual Environments: Results of a Three-Condition Within-Subjects Study With Long Intersession Intervals. *Virtual Reality*, 24(4):783–796.

Porter, L., Bailey-Lee, C., and Simon, B. (2013). Halving Fail Rates Using Peer Instruction: A Study of Four Computer Science Courses. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE)*, pages 177–182.

Potts, S. R., McCuddy, W. T., Jayan, D., and Porcelli, A. J. (2019). To Trust, or Not to Trust? Individual Differences in Physiological Reactivity Predict Trust Under Acute Stress. *Psychoneuroendocrinology*, 100:75–84.

Prada, R. and Rato, D. (2022). Socially Interactive Agents in Games. In Lugrin, B., Pelachaud, C., and Traum, D., editors, *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, pages 493–526. Association for Computing Machinery, first edition.

Premack, D. and Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind? *Behavioral and Brain Sciences*, 1(4):515–526.

Prochazkova, E., Prochazkova, L., Giffin, M. R., Scholte, H. S., De Dreu, C. K. W., and Kret, M. E. (2018). Pupil Mimicry Promotes Trust Through the Theory-of-Mind Network. *Proceedings of the National Academy of Sciences*, 115(31):E7265–E7274.

Punyatoya, P. (2019). Effects of Cognitive and Affective Trust on Online Customer Behavior. *Marketing Intelligence & Planning*, 37(1):80–96.

Putze, S., Alexandrovsky, D., Putze, F., Höffner, S., Smeddinck, J. D., and Malaka, R. (2020). Breaking the Experience: Effects of Questionnaires in VR User Studies. In *Proceedings of the 38th CHI Conference on Human Factors in Computing Systems*.

Quesque, F. and Rossetti, Y. (2020). What Do Theory-of-Mind Tasks Actually Measure? Theory and Practice. *Perspectives on Psychological Science*, 15(2):384–396.

Raffard, S., Salesse, R. N., Bortolon, C., Bardy, B. G., Henriques, J., Marin, L., Stricker, D., and Capdevielle, D. (2018). Using Mimicry of Body Movements by a Virtual Agent to Increase Synchronization Behavior and Rapport in Individuals With Schizophrenia. *Scientific Reports*, 8.

Randhavane, T., Bera, A., Kapsaskis, K., Sheth, R., Gray, K., and Manocha, D. (2019). Eva: Generating Emotional Behavior of Virtual Agents Using Expressive Features of Gait and Gaze. In *Proceedigns of the 8th ACM Symposium on Applied Perception (SAP)*.

Ranjbartabar, H., Richards, D., Bilgin, A. A., and Kutay, C. (2019). First Impressions Count! The Role of the Human's Emotional State on Rapport Established With an Empathic Versus Neutral Virtual Therapist. *IEEE Transactions on Affective Computing*, 12(3):788–800.

Reinhard, R., Telatar, E., and Humayoun, S. R. (2022). Comparison of Object Detection in Head-Mounted and Desktop Displays for Congruent and Incongruent Environments. *Big Data and Cognitive Computing*, 6(1):28.

Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in Close Relationships. *Journal of Personality and Social Psychology*, 49(1):95–112.

Riley, V. (1996). Operator Reliance on Automation: Theory and Data. In Parasuraman, R. and Mouloua, M., editors, *Automation and Human Performance: Theory and Applications*, pages 19–35. Taylor & Francis.

Ring, L., Utami, D., and Bickmore, T. W. (2014). The Right Agent for the Job? In *Proceedings of the 14th International Conference on Intelligent Virtual Agents (IVA)*, pages 374–384.

Robertson, S., Solomon, R., Riedl, M., Gillespie, T. W., Chociemski, T., Master, V., and Mohan, A. (2015). The Visual Design and Implementation of an Embodied Conversational Agent in a Shared Decision-Making Context (eCoach). In *Proceedings of the 2nd International Conference on Learning and Collaboration Technologies (LCT)*, pages 427–437.

Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., and Doerr, C. E. (2008). Interface Agents as Social Models for Female Students: The Effects of Agent Visual Presence and Appearance on Female Students' Attitudes and Beliefs. *Computers in Human Behavior*, 24(6):2741–2756.

Ross, S. A. (1973). The Economic Theory of Agency: The Principal's Problem. *The American Economic Review*, 63(2):134–139.

Rotenberg, K. J., Petrocchi, S., Lecciso, F., and Marchetti, A. (2015). The Relation Between Children's Trust Beliefs and Theory of Mind Abilities. *Infant and Child Development*, 24(2):206–214.

Rotter, J. B. (1967). A New Scale for the Measurement of Interpersonal Trust. *Journal of Personality*, 35(4):651–665.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. (1998). Not So Different After All: A Cross-Discipline View of Trust. *Academy of Management Review*, 23(3):393–404.

Ruocco, M., Mou, W., Cangelosi, A., Jay, C., and Zanatto, D. (2021). Theory of Mind Improves Human's Trust in an Iterative Human-Robot Game. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI)*, pages 227–234.

Russell, P. A., Hosie, J. A., Gray, C. D., Scott, C., Hunter, N., Banks, J. S., and Macaulay, M. C. (1998). The Development of Theory of Mind in Deaf Children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39(6):903–910.

Sakuma, H., Takahashi, H., Ogawa, K., and Ishiguro, H. (2023). Immersive Role-Playing With Avatars Leads to Adoption of Others' Personalities. *Frontiers in Virtual Reality*, 4.

Santhoshkumar, R. and Geetha, M. K. (2019). Deep Learning Approach for Emotion Recognition From Human Body Movements With Feedforward Deep Convolution Neural Networks. *Procedia Computer Science*, 152:158–165.

Saxe, R. (2009). Theory of Mind (Neural Basis). *Encyclopedia of Consciousness*, 2:401–410.

Schroeder, N. L., Adesope, O. O., and Gilbert, R. B. (2013). How Effective Are Pedagogical Agents for Learning? A Meta-Analytic Review. *Journal of Educational Computing Research*, 49(1):1–39.

Sengul, M., Gimeno, J., and Dial, J. (2012). Strategic Delegation: A Review, Theoretical Integration, and Research Agenda. *Journal of Management*, 38(1):375–414.

Shapiro, S. P. (2005). Agency Theory. *Annual Review of Sociology*, 31:263–284.

Simon, S. C. and Greitemeyer, T. (2019). The Impact of Immersion on the Perception of Pornography: A Virtual Reality Study. *Computers in Human Behavior*, 93:141–148.

Sin, D., Chew, T., Chia, T. K., Ser, J. S., Sayampanathan, A., and Koh, G. (2019). Evaluation of Constructing Care Collaboration-Nurturing Empathy and Peer-to-Peer Learning in Medical Students Who Participate in Voluntary Structured Service Learning Programmes for Migrant Workers. *BMC Medical Education*, 19.

Slater, M. (2003). A Note on Presence Terminology. *Presence Connect*, 3(3):1–5.

Slater, M. (2009). Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557.

Slater, M. and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence (Camb)*, 6(6):603–616.

Song, S. W. and Shin, M. (2022). Uncanny Valley Effects on Chatbot Trust, Purchase Intention, and Adoption Intention in the Context of E-Commerce: The Moderating Role of Avatar Familiarity. *International Journal of Human-Computer Interaction*.

Sousa Santos, B., Dias, P., Pimentel, A., Baggerman, J.-W., Ferreira, C., Silva, S., and Madeira, J. (2008). Head-Mounted Display Versus Desktop for 3D Navigation in Virtual Reality: A User Study. *Multimedia Tools and Applications*, 41(1):161–181.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., and Waters, K. (1996). When the Interface Is a Face. *Human-Computer Interaction*, 11(2):97–124.

Steffel, M. and Williams, E. F. (2018). Delegating Decisions: Recruiting Others to Make Choices We Might Regret. *Journal of Consumer Research*, 44(5):1015–1032.

Steffel, M., Williams, E. F., and Perrmann-Graham, J. (2016). Passing the Buck: Delegating Choices to Others to Avoid Responsibility and Blame. *Organizational Behavior and Human Decision Processes*, 135:32–44.

Stein, J.-P. and Ohler, P. (2017). Venturing Into the Uncanny Valley of Mind – the Influence of Mind Attribution on the Acceptance of Human-Like Characters in a Virtual Reality Setting. *Cognition*, 160:43–50.

Stout, N., Dennis, A. R., and Wells, T. M. (2014). The Buck Stops There: The Impact of Perceived Accountability and Control on the Intention to Delegate to Software Agents. *AIS Transactions on Human-Computer Interaction*, 6(1):1–15.

Straßmann, C. and Krämer, N. C. (2017). A Categorization of Virtual Agent Appearances and a Qualitative Study on Age-Related User Preferences. In *Proceedings of the 17th International Conference on Intelligent Virtual Agents (IVA)*, pages 413–422.

Straßmann, C. and Krämer, N. C. (2018). A Two-Study Approach to Explore the Effect of User Characteristics on Users' Perception and Evaluation of a Virtual Assistant's Appearance. *Multimodal Technologies and Interaction*, 2(4).

Sturgeon, S., Palmer, A., Blankenburg, J., and Feil-Seifer, D. (2019). Perception of Social Intelligence in Robots Performing False-Belief Tasks. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*.

Sun, N. and Botev, J. (2021a). Intelligent Autonomous Agents and Trust in Virtual Reality. *Computers in Human Behavior Reports*, 4:100146.

Sun, N. and Botev, J. (2021b). Virtual Agent Representation for Critical Transactions. In *Proceedings of the 13th International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE)*, pages 25–29.

Sun, N. and Botev, J. (2021c). Why Do We Delegate to Intelligent Virtual Agents? Influencing Factors on Delegation Decisions. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI)*, pages 386–390.

Sun, N. and Botev, J. (2023). Technological Immersion and Delegation to Virtual Agents. *Multimodal Technologies and Interaction*, 7(11):106.

Sun, N., Botev, J., Khaluf, Y., and Simoens, P. (2022). Theory of Mind and Delegation to Robotic Virtual Agents. In *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 454–460.

Sun, N., Botev, J., and Simoens, P. (2023). The Effect of Rapport on Delegation to Virtual Agents. In *Proceedings of the 23rd International Conference on Intelligent Virtual Agents (IVA)*.

Sundar, S. S. and Kim, J. (2019). Machine Heuristic: When We Trust Computers More Than Humans With Our Personal Information. In *Proceedings of the 37th CHI Conference on Human Factors in Computing Systems*.

Sundström, G. A. and Salvador, A. C. (1994). Cooperative Human-Computer Decision Making: An Experiment and Some Design Implications. In *Proceedings of the 38th Human Factors and Ergonomics Society Annual Meeting (HFES)*, pages 220–224.

Surprenant, A. M. (2012). *Measuring Trust In Virtual Worlds: Avatar-Mediated Self-Disclosure*. PhD thesis, University of Central Florida.

Takagi, H. and Terada, K. (2021). The Effect of Anime Character's Facial Expressions and Eye Blinking on Donation Behavior. *Scientific Reports*, 11.

Tamborini, R., Novotny, E., Prabhu, S., Hofer, M., Grall, C., Klebig, B., Hahn, L. S., Slaker, J., Ratan, R. A., and Bente, G. (2018). The Effect of Behavioral Synchrony With Black or White Virtual Agents on Outgroup Trust. *Computers in Human Behavior*, 83:176–183.

Tanaka, H., Saga, T., Iwauchi, K., and Nakamura, S. (2023). Acceptability and Trustworthiness of Virtual Agents by Effects of Theory of Mind and Social Skills Training. In *Proceedings of the 17th International Conference on Automatic Face and Gesture Recognition (IEEE FG)*.

Tarnowski, P., Kolodziej, M., Majkowski, A., and Rak, R. J. (2017). Emotion Recognition Using Facial Expressions. *Procedia Computer Science*, 108:1175–1184.

Tarr, B., Slater, M., and Cohen, E. (2018). Synchrony and Social Connection in Immersive Virtual Reality. *Scientific Reports*, 8.

Thellman, S., Silvervarg, A., Gulz, A., and Ziemke, T. (2016). Physical vs. Virtual Agent Embodiment and Effects on Social Interaction. In *Proceedings of the 16th International Conference on Intelligent Virtual Agents (IVA)*, pages 412–415.

Theonas, G., Hobbs, D., and Rigas, D. (2008). Employing Virtual Lecturers' Facial Expressions in Virtual Educational Environments. *The International Journal of Virtual Reality*, 7(1):31–44.

Thielsch, M. T., Meeßen, S. M., and Hertel, G. (2018). Trust and Distrust in Information Systems at the Workplace. *PeerJ*, 6:e5483.

Tickle-Degnen, L. and Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, 1(4):285–293.

Torre, I., Carrigan, E., McDonnell, R., Domijan, K., McCabe, K., and Harte, N. (2019). The Effect of Multimodal Emotional Expression and Agent Appearance on Trust in Human-Agent Interaction. In *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games*.

Torre, I., Goslin, J., and White, L. (2020). If Your Device Could Smile: People Trust Happy-Sounding Artificial Agents More. *Computers in Human Behavior*, 105:106215.

van Wissen, A., Vinkers, C., and van Halteren, A. (2016). Developing a Virtual Coach for Chronic Patients: A User Study on the Impact of Similarity, Familiarity and Realism. In *Proceedings of the 11th International Conference on Persuasive Technology (PERSUASIVE)*, pages 263–275.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3):425–478.

Verberne, F. M. F., Ham, J. R. C., and Midden, C. J. H. (2012). Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors*, 54(5):799–810.

Verberne, F. M. F., Ham, J. R. C., and Midden, C. J. H. (2014). Familiar Faces: Trust in Facially Similar Agents. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Verberne, F. M. F., Ham, J. R. C., and Midden, C. J. H. (2015). Trusting a Virtual Driver That Looks, Acts, and Thinks Like You. *Human Factors*, 57(5):895–909.

Verosky, S. C. and Todorov, A. (2010). Differential Neural Responses to Faces Physically Similar to the Self as a Function of Their Valence. *NeuroImage*, 49(2):1690–1698.

Watson, A. C., Nixon, C. L., Wilson, A., and Capage, L. (1999). Social Interaction Skills and Theory of Mind in Young Children. *Developmental Psychology*, 35(2):386–391.

Waytz, A., Heafner, J., and Epley, N. (2014). The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle. *Journal of Experimental Social Psychology*, 52:113–117.

Weimann, G. (1999). *Communicating Unreality: Modern Media and the Reconstruction of Reality*. Sage Publications.

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., and André, E. (2019). "Do You Trust Me?" Increasing User-Trust by Integrating Virtual Agents in Explainable AI Interaction Design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA)*, pages 7–9.

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth About False Belief. *Child Development*, 72(3):655–684.

Willis, M. L., Palermo, R., and Burke, D. (2011). Social Judgments Are Influenced by Both Facial Expression and Direction of Eye Gaze. *Social Cognition*, 29(4):415–419.

Wilson, R. (1968). The Theory of Syndicates. *Econometrica*, 36(1):119–132.

Wiltermuth, S. S. and Heath, C. (2009). Synchrony and Cooperation. *Psychological Science*, 20(1):1–5.

Wong, J. W. and McGee, K. (2012). Frown More, Talk More: Effects of Facial Expressions in Establishing Conversational Rapport With Virtual Agents. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents (IVA)*, pages 419–425.

Wyland, C. L. and Forgas, J. P. (2010). Here's Looking at You Kid: Mood Effects on Processing Eye Gaze as a Heuristic Cue. *Social Cognition*, 28(1):133–144.

Yasavur, U., Lisetti, C., and Rishe, N. (2014). Let's Talk! Speaking Virtual Counselor Offers You a Brief Intervention. *Journal on Multimodal User Interfaces*, 8(4):381–398.

Yee, N. and Bailenson, J. (2007). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, 33(3):271–290.

Yoshida, W., Dolan, R. J., and Friston, K. J. (2008). Game Theory of Mind. *PLoS Computational Biology*, 4(12).

Yukl, G. and Fu, P. P. (1999). Determinants of Delegation and Consultation by Managers. *Journal of Organizational Behavior*, 20(2):219–232.

Yuksel, B. F., Collisson, P., and Czerwinski, M. (2017). Brains or Beauty: How to Engender Trust in User-Agent Interactions. *ACM Transactions on Internet Technology*, 17(1):1–20.

Zhao, Y., Ham, J. R. C., and van der Vlist, J. (2018). Persuasive Virtual Touch: The Effect of Artificial Social Touch on Shopping Behavior in Virtual Reality. In *Proceedings of the 6th International Workshop on Symbiotic Interaction*, pages 98–109.

Zhou, M. X., Mark, G., Li, J., and Yang, H. (2019). Trusting Virtual Agents: The Effect of Personality. *ACM Transactions on Interactive Intelligent Systems*, 9(10):1–36.

Zhou, S., Bickmore, T. W., Paasche-Orlow, M. K., and Jack, B. (2014). Agent-User Concordance and Satisfaction With a Virtual Hospital Discharge Nurse.

In *Proceedings of the 14th International Conference on Intelligent Virtual Agents (IVA)*, pages 528–541.