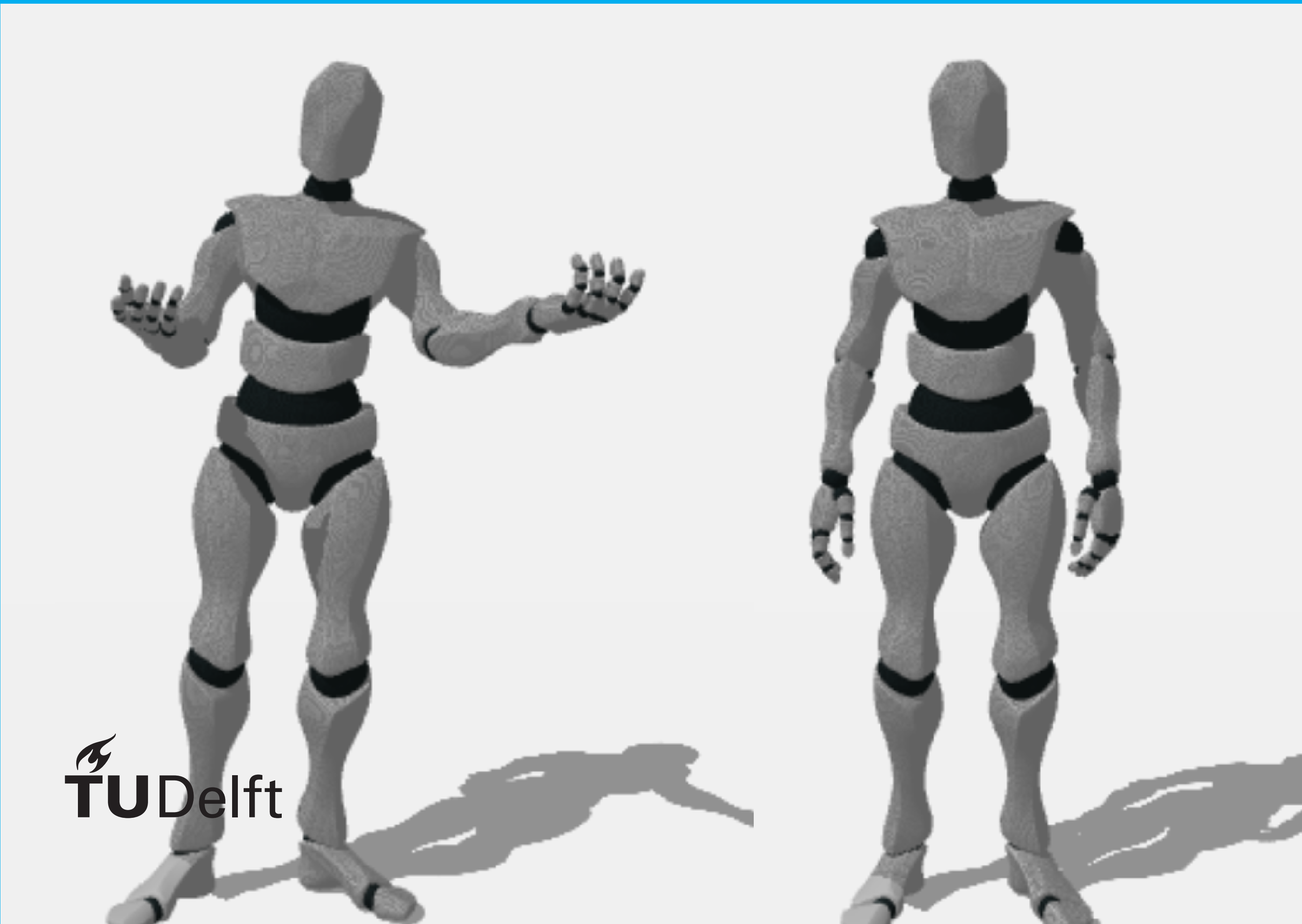
Exploring the effect of gestures on perceived integrity in human-agent interaction

Alan van Rossum



Exploring the effect of gestures on perceived integrity in human-agent interaction

by

Alan van Rossum

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on August 30, 2022

Student number:	4293932	
Project duration:	April 19, 2021 – August 30, 2022	
Thesis committee:	Prof. dr. C. Jonker,	TU Delft, thesis advisor
	Dr. M. L. Tielman,	TU Delft, supervisor
	Msc. S. Mehrotra,	TU Delft, daily supervisor
	Dr. U. Gadiraju,	TU Delft, external member

This thesis is confidential and cannot be made public until December 31, 2022.

Preface

I would like to thank my supervisors; Myrthe Tielman and Siddharth Mehrotra, for their continued supervision. Their patience and ability to think along with me were essential to the completion of this research. I also want to thank Catholijn Jonker and Ujwal Gadiraju for their roles as thesis committee members, Pei-Yu Chen and Mohammed Al Owayyed for their valuable feedback on this report, and all participants of the experiment for their time and effort. Finally, a special thanks to Eliana Tal for being my number one support.

Alan van Rossum
Delft, July 2022

Contents

1 Introduction

Cracking the enigma code, landing on the moon, creating the internet. Mankind's greatest achievements of the last century have all been the result of human-computer collaboration and teamwork. It can be easy for developers and engineers to trust their own creations enough to cooperate with them, but the same can't be said for the rest of the world. Smart speakers like Alexa for example are not trusted enough not to spy on their owners, and are as a result turned off or not bought at all [1]. In order for developers to make software that people use, they have to know how to make software that people trust. The initial goal of this research was to help developers do exactly that.

The first step in this research is defining trust. Trust between two parties relies on the characteristics of both the trustor and the trustee [2]. The trusting party needs to have a certain propensity to trust, or no form of trust can take place. The trustee needs to possess some level of trustworthiness. To understand what makes someone trustworthy, I look at the social psychology definition by Mayer (1995) who describes the foundation of trustworthiness as consisting of three factors: ability, benevolence and integrity [2]. Although these characteristics can describe a persons trustworthiness, how these factors are perceived by others is what determines their perceived trustworthiness. Thus, one could increase their perceived ability by performing better, or by presenting themselves differently, for example by appearing cold or using numerical rhetoric [3], [4]. Similarly, one's perceived benevolence could be increased by being more kind and thoughtful, but also by hiding your egocentric side and showcasing your altruistic side. Lastly, ones perceived integrity can be increased by being honest and consistent, or by changing your presented integrity. This could for example be done by using hand gestures while speaking or even lying in specific situations [5], [6]. Choshen (2020) showed for example that people lie about unlikely scenarios to avoid looking like liars, and that speaking the truth in such a scenario would indeed have lowered their perceived honesty [7].

Besides being important for building interpersonal relationships, trust is also required for the acceptance of technology [8]. For this reason it is interesting for developers of computers that interact with humans to find ways to increase the perceived trustworthiness of their systems. One problem here is that humans tend to respond to technology socially, but most technology does not have access to the social signals expressed through the face, body, voice and motion like humans do [9]. The solution to this problem could lie in the embodiment of technological systems, because when technology takes the form of a social actor like a robot or a virtual embodied agent, they have access to extra characteristics like voice, facial animations and gestures [10]. But can proper use of these extra features influence an embodied agent's perceived trustworthiness in the same way they do for humans? This is one thing that researchers in the field of human-agent interaction (HAI) are trying to find out. They found that making good use of these features can for example increase sense of collaboration and perceptions of social presence [11], [12]. Other research has explored the effect of embodiment on trusting beliefs. For the ability component of the perceived trustworthiness, research has indeed found that the embodiment of agents and the use of non-verbal cues can improve perceptions of the ability of that agent [13]. However, the effect of tone of voice and gestures by embodied agents on perceived integrity and benevolence is something that has not been extensively studied. Even though the effect of gestures and speech on perceived integrity has been studied with humans and was found to affect perceived integrity, research regarding the use of gestures by embodied agents is lacking [6], [14]. I had to decide which feature of embodiment to explore; gestures, facial animations, posture. Since psychologists attribute communication between humans to rely for over 50% on body language and because gestures have been shown to improve perceptions of honesty in humans, I choose to explore the effect of body language in the form of gestures on perceived integrity between humans and agents [15]. This leads to the following research question:

- How can an agent's integrity presented through gestures influence the belief of that agent's integrity in a human?

The main reason why answering this question is important is because it can help in facilitating the design of agents that aim to gain users trust in HAI, ultimately aiming to avoid disuse and increase adoption [16]. Additionally, this research can be used as a stepping stone for future work that further explores the effects of different kinds of agents on ability and benevolence.

To answer the aforesaid research question, I performed a user study that compares the perceived integrity of a virtual agent that uses gestures with one that doesn't. Before I could start such an experiment, I needed to answer two questions by exploring related work. This has been done in chapter 2.

1. How do I define the role of integrity as a part of trust in human-agent interaction?
2. How do humans use gestures to alter their perceived integrity?

Answering these questions provided the necessary knowledge to conduct an experiment that measured the effect of gestures on perceived integrity in human-agent interaction.

2 Foundation

The first goal of this chapter is to obtain concrete definitions of trust and integrity. This will be done by first exploring the definitions of these terms in inter human trust, and then exploring their definitions in the field of HCI. Since definitions may vary between sources, it is necessary to compare them and ultimately outline my own definitions in order to have a strong grounding for the rest of this research and to make any useful conclusions. The definition of integrity needs to be implementable in a way that a participants' perceived integrity of an agent can be strictly measured.

The second goal of this chapter is to explore different ways an agent could use gestures to present itself differently. I do this by looking at the use of gestures by both humans and embodied agents, and how this use of gestures affects trusting beliefs.

By defining trust and integrity and by exploring the use of gestures, I will have all the tools available to create an experiment that can measure the effect of gestures on the perceived integrity of virtual embodied agents.

2.1. Defining Trust

As said by Husted (1989), "The definition of trust is problematic because there are such a wide variety of approaches to the concept" [17]. The definition of trust is different in different contexts. The field of HCI is at its core interdisciplinary. It involves the study of humans, as well as the study of computers. For this reason, I will be looking at definitions of trust from these different perspectives. For the human side, I will look at notions of trust in contexts of philosophy and psychology. For the computer side, I will look at notions of trust in the field of computer science.

2.1.1. Trust between humans

Held (1968) reviewed several ways of looking at trust [18]. Where other philosophers like Tullock (1967) consider trust being equal to the ability to accurately predict someone's behaviour, Held concludes the opposite [19]. No trust is needed when you are certain of another persons actions. It is exactly in uncertain situations that trust is required to reach outcomes that benefit everyone, like in the famous prisoner's dilemma. This uncertainty also leads to vulnerability. Allowing yourself to be vulnerable to the uncertain actions of another is exactly what trusting that person is. Even though philosophers tend to not always agree on the same definitions, according to the Stanford Encyclopedia of Philosophy there appears to be at least some consensus that vulnerability and reliance on others are the amongst the basis of trust [20]. This notion of vulnerability is important: there is no need for trust in a situation where no form of betrayal of trust can take place.

Besides philosophy, trust has also been studies in the field of management and entrepreneurship. An important work in this area of trust is that of Mayer (1995) [2]. This work suggests a general trust model comprised of ability, benevolence and integrity. They define ability as skills and competencies that allow an entity to have influence in a specific domain. Benevolence is defined as the extent to which an entity is believed to act non-egocentrically, to benefit their trustors. Finally integrity is defined as adhering to principles that are acceptable to the trustor. It should be mentioned that this work was founded to measure trust in an organisational setting, and it provides a framework of measuring employees' perception of the trustworthiness of their employer. Adaptations of this framework have also been used in trust between teams, as well as between individuals [21], [22].

Some important things to know about trust have to do with time and context. Berg (1995) explains the notion that trust is built up over time [23]. Initial beliefs of trustworthiness are often based on reputation appearance, but these beliefs may change after longer or multiple interactions. In addition to this, trust is dependent on context. Beliefs of trustworthiness are subjective rather than objective, and are dependant on the trustor [24]. This means that in order to correctly assess the trustworthiness of an entity, multiple people (to counter subjectivity) need to have multiple interactions with that entity.

2.1.2. Trust in HCI

In the field of HCI research has been done on agents as social actors and the role of trust in interacting with them. Some research personifies agents and evaluate them in the same way as humans, based on the ABI model [25], [26]. The earlier mentioned influence of the amount of interactions on trust is also explored with agents, for example by Carter (2003) [27]. They found that when humans interacted with agents over a longer period of time, their beliefs about the agents ability, benevolence and integrity eventually converged to how the agent was intended.

2.2. Integrity

Having integrity can be defined in many ways: fairness, honesty, openness, consistency, predictability, reliability, dependability, discreetness, communicating clearly, keeping promises and acting according to your own principles, to name a few. This big list of synonyms helps with getting a broad idea of the meaning of the word. However, to help in formulating a more concrete definition that can be used in a HAI experiment, other works that measure or manipulate integrity in the fields of philosophy, psychology and HCI have to be explored.

2.2.1. Integrity between humans

A literature review by Palanski (2007) provides an overview of relevant integrity definitions in philosophy [5]. They outline five general categories of integrity: wholeness, consistency of words and actions, consistency in adversity, being true to one-self, and moral/ethical behaviour. Wholeness describes integrity as relating to a person as a whole, rather than to specific aspects of that person. This tells us that in order to measure integrity, we need to measure integrity of the person or agent as a whole, and we can not look at certain actions and determine the integrity of that action specifically. Being true to one-self is seen as acting according to your own values, rather than being influenced by exterior motives. Interestingly, the fifth category describes moral or ethical behaviour. McFall (1987) states that acting according to your own immoral principle can be seen as still having personal integrity since you are at least consistent [28]. Palanski (2007) summarises that a majority of moral philosophers say that these principles need to be moral or ethical in order for the person to have integrity [5].

For this research it will be important to measure the beliefs of humans about the integrity of agents. Mayer (1995) provides integrity definitions used to measure integrity in social interaction between employees and management [2]. The main components of their definition are keeping promises, being fair in dealings with others, being consistent, and acting according to sound principles and values. Other research in human-human interaction measures integrity by simply 'being honest' and 'having integrity' [29].

2.2.2. Integrity in HCI

Many works in HCI or HRI (human-robot interaction) personify the digital agent and apply the same definitions of integrity to them as they would to humans. For example, McKnight (2002) combines the traditional trusting beliefs of competence, benevolence and integrity with trusting intentions, defined as the decision to make oneself vulnerable to the trustee, in their case an online vendor [26]. They defined the integrity part of these beliefs as honesty and promise keeping. Jensen (2018) measures the integrity of a drone system as being truthful in communication, honest, keeping commitments, being sincere and genuine, and performing as expected [30]. Other research in the field of integrity as moral motivator in economic interactions simply defines integrity as 'not lying' [31].

2.2.3. My definition

Many definitions of integrity that I found in studies relating to trust in humans as well as agents have mentioned honesty as a central part of the definition [29], [2], [31], [32], [26]. Additionally, honesty is something that could be more objectively coded into an agent than more vague definitions like wholeness. For these two reasons I have decided to reduce the definition of integrity to honesty for the design of experiment. This allows any agents to be quantitatively compared by their amount of lying.

Although being honest or dishonest can still be subjective, for an isolated interaction like a trading game it can be objectively defined by the metrics of the system [33]. One could for example create an agent that is programmed to lie in 20 percent of their turns. This would make them have objectively more honest than an agent that lies in 50 percent of their turns, provided all other factors stay the same. To measure integrity, two Likert scales on integrity will be used [34]. The first scales are the traditional scale used by Mayer (1995), as well as the more recent scale with a fcus on technology rather than humans by Jensen (2018) [2], [30]. Two statements about honesty from the Jensen scale are used to measure perceived honesty specifically.

2.3. Gestures

For a developer to design an agent that aims to gain the users trust, they have to know the effect of certain design choices on perceptions of trustworthiness. For unembodied agents, these design choices have to do with the capabilities and functionality of the agent. Embodied agents can also have access to voice, facial animations and gestures. Body language accounts for 55% of the communication between humans [15]. Since gestures are a big part of body language and because body language plays a major part in communication, I want to explore the effect of gestures on perceptions of integrity between humans and agents. To know whether gestures could alter perceived integrity in agents I look at the use of gestures by humans and their effects on perceptions of integrity and honesty.

One study that examined the behaviour of parole clients showed that their nonverbal behaviours affected how honest they were perceived, which in turn actually led to different decisions by their parole officers [14]. This shows that the use of gestures can affect perceived honesty. But what kind of gestures affect integrity specifically? One study found that the gesture of a hand over the heart (figure 2.1) increased the level of honesty perceived by other humans [6]. The body language academy describes another gesture of showing hands with open palms (figure 2.2) as being associated with trust, honesty and submission [35]. Finally, While administering an oath (figure 2.3, the most common gesture is to raise the right hand [36]. The swearing of an oath is associated with telling the absolute truth, and thus being completely honest.

Although these specific gestures have been studied when used by humans, no similar studies were found that measure how perceptions of honesty are affected by the use of such gestures by embodied agents. There are studies that show that human perceptions of trust are indeed affected by the visual appearance of embodied agents [37]. Additionally, the use of gestures by agents does affect some perceptions like likeability [38]. For these reasons I hypothesise that the use of gestures by embodied agents can increase how honest they are perceived by humans.

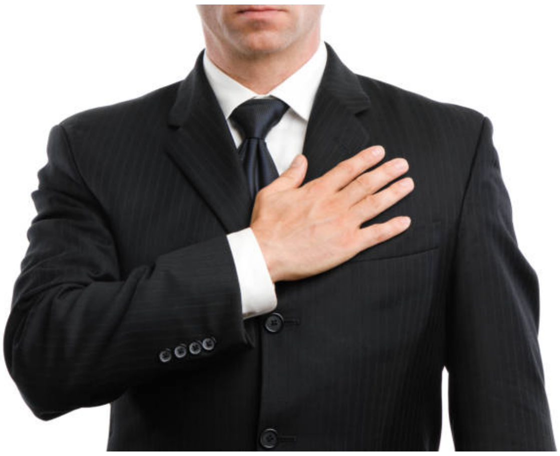


Figure 2.1: Hand over the heart

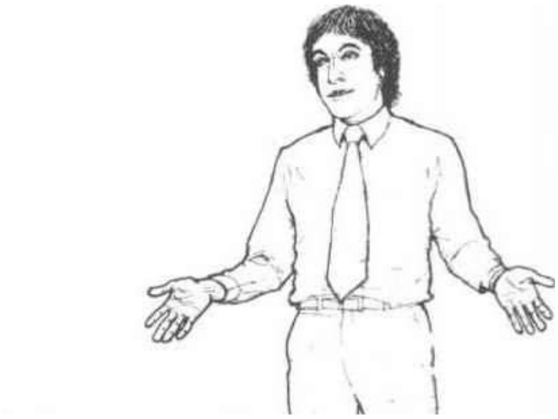


Figure 2.2: Open palms

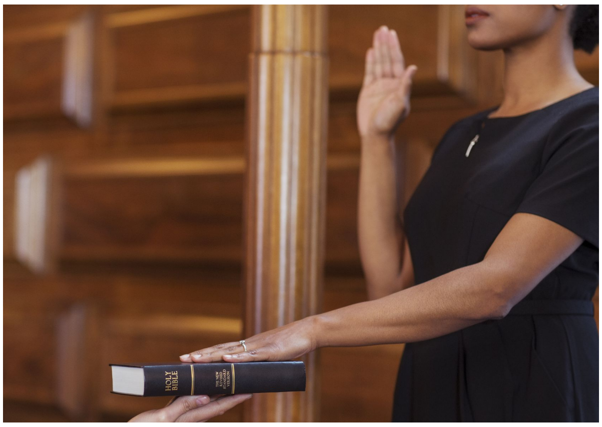


Figure 2.3: Swearing an oath

3 Specification

The goal of this research is to measure the effect of the use of gestures by agents on their perceived integrity by humans. In order to conduct an experiment that finds an answer to the research question, a HAI scenario needs to be designed and developed. This scenario must adhere to specific requirements in order to generate useful and accurate results.

3.1. Requirements

Since we are looking at trust in agents, there needs to be a reason for the human to trust the agent(s). Therefore, there needs to be some form of cooperation between the human and the agent(s): the human should require the agents help in some form or way. There can be no trust without vulnerability (2.1.1). Therefore, the human needs to be rely on the agent in some form, being vulnerable to the results of their actions. Since trust builds up and converges over time, there have to be multiple interactions between the human and the agent(s) (2.1.1). Since it is the effect of gestures that we want to measure, the agent(s) in question has to be embodied and able to use gestures (2.3). Finally, there should be some form of measuring honesty during the scenario. Even though a questionnaire on perceived integrity can be used to measure human perceptions after any scenario, this only measures the conscious perceptions. Any subconscious decisions can only be revealed by measuring what humans actually do while interacting with the agent(s).

1. The scenario must be cooperative
2. The scenario must feature human vulnerability
3. There must be multiple interactions between the human and the agent(s)
4. The agent(s) must be embodied.
5. The agent(s) must be able to use gestures.
6. The scenario should facilitate measurements for perceptions of integrity

3.2. Scenario

A major decision that needs to be made for the scenario is the amount of agents involved. In a single-agent setting the focus of the human is undivided. The user is however given no choice on who they want to interact with. On the contrary, a multi-agent setting that allows the human more freedom of choice. By choosing who to interact the human can provide us with valuable information about who the human trusts more at a certain point during the interaction. A multi-agent interaction with individually different agents can allow for an easier comparison between different types of agents and the effect on the users beliefs about them. After brainstorming multiple scenario ideas, which can be viewed in appendix F together with their pros and cons, I finally settled on one: the agent-assisted trivia quiz.

3.2.1. Agent-assisted trivia quiz

This scenario features a human user with as goal to correctly answer multiple-choice quiz questions asked by the computer. The user has access to the assistance of two embodied virtual agents. Since the quiz has to adhere to the requirements 3.1, certain design decisions were made for this scenario. First of all, the quiz is cooperative. The user aims to obtain as many points as possible by answering questions correctly. The agents aim to assist the user by providing him with hints to answer the questions. The agents have a personal goal of being picked as many times as possible, and the user helps them to reach this goal by allowing them to help him. To make sure the user has some vulnerability in this scenario, the hints given by the agents will not always be correct, and they can only choose one agent for each question. This means that the user has the risk of giving incorrect answers if they choose to blindly trust whatever the agents say. Additionally, asking for a hint will cost the user a point. To also increase the risk of not trusting any of the agents, the questions will be very hard and answering a question incorrectly will cost the user two points. The quiz features 16 questions, and for each question the user has a new choice on who they want to trust for this round. This ensures that trust can be built up over these multiple interactions.

Since I want to measure the effect of gestures on user beliefs, the agents are embodied. They have access to multiple gestures that they can employ while speaking. The agents are designed to only differ in one factor: their use of gestures. One of the agents uses gestures to appear more honest, while the other agent remains in their idle animation for the entire duration of the quiz. Both agents feature a neutral grey look, and their position and speaking order, left or right, is randomised for each user. This makes sure that if a difference is found between the users' beliefs in both agents, that this difference has to be based on their use of gestures. Lastly, both agents are not completely honest. They can only give correct hints to 75% of the questions, but they propose to the user in each round as if they know everything. This ensures that a user that starts by only trusting one agent has a reason to switch their trust to the other one. Once this other agent eventually makes a mistake, the user will have to come up with a strategy that they think will provide them with the most points. While some users may choose to try to answer the hard questions on their own, I hope that most users will let their 'feeling' tell them who is more honest. If users do this, I know that their decision can only be made on the agents use of gestures, either consciously or subconsciously, since all other factors are identical.

An exact walkthrough and other technical details of the implementation of the agent-assisted trivia quiz can be found in the next chapter.

4 Implementation

The goal of this chapter is to provide insight into the implementation of the Agent-Assisted Trivia Quiz. This chapter will discuss two aspects of the implementation. First, I will cover the quiz walkthrough. This walkthrough describes exactly what the user experiences when doing the quiz. This section can help in understanding the point of view of the user, and show what kind of decisions they have to make while playing the game. Additionally, the in-depth description of the interfaces improves the reproducibility of this research. After this walkthrough I will discuss the technical details of the implementation. This section explains how the quiz was programmed and certain design decisions that improve usability and user privacy.

4.1. Quiz walkthrough

4.1.1. Tutorial and instructions

The tutorial teaches the mechanics of the game to the user. It explains which buttons to click to play through the game, and explains the instructions of the rounds and points system, explaining that there are 16 rounds, 5 points are gained for a correct answer, 2 are lost for a wrong answer and 1 is lost for choosing to retrieve the hint of an agent. It finishes with a practice question which is not counted towards the final score. An exact rundown of the tutorial and the instructions provided can be viewed in appendix G.

4.1.2. Questions

For illustration, figure 4.1 shows the diagram of how the user progresses through the quiz. Each state and state transition is also explained in words in this section.

0. **Initialise and new question**. This is the state that the user enters after finishing the tutorial. If the user enters this state for the first time, their points are set to 10. At random either the left or the right agent is designated as the gesturing agent. The other agent is designated as the idle agent, i.e. an agent that does speak, but does not use any gestures besides an idle animation. The first question is presented on the screen together with all four multiple choice answers. All 16 questions can be viewed in appendix C. The user needs to listen to the proposals of both agents first. After reading the question, the user can press the start button to proceed to state 1.

1. **Randomising speaking order.** In this state, an agent is selected at random to speak first. This random selection leads to a state 2 that either starts with the left agent or the right agent speaking.

2. **Agent proposals.** The first agent to speak starts their proposal. This causes the area around the agent to light up green, and makes the text that is being spoken appear below the agent. If this agent is the gesturing agent, they will use one of three gestures viewable in figure 5.1. If the agent is the idle agent, they will use the idle animation from figure 5.1. After an agent finishes speaking, the area around the agent is set back to white, their animation changes back to idle, and the text below the agent is removed. After the first agent finishes speaking, the second agent starts their speech. After both agents finish speaking, the game changes to state 3.

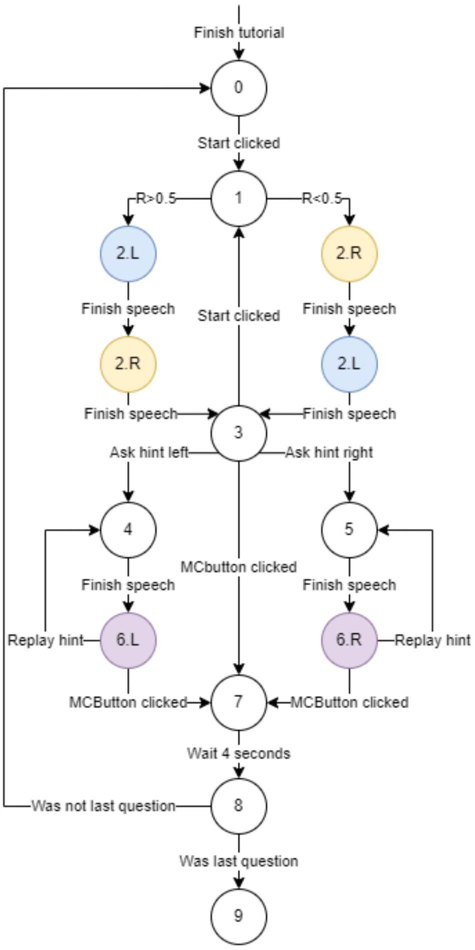


Figure 4.1: State diagram

3. **Player choice.** In this state, the user can choose their action. The multiple choice buttons are activated, so the player can choose to answer the question by clicking one of them. This will take them to state 7. Other options for the player are choosing to hear what one of the two agents has to say to help them. Choosing an agent costs one point. Choosing the left agent takes the user to state 4. Choosing the right agent takes the user to state 5.

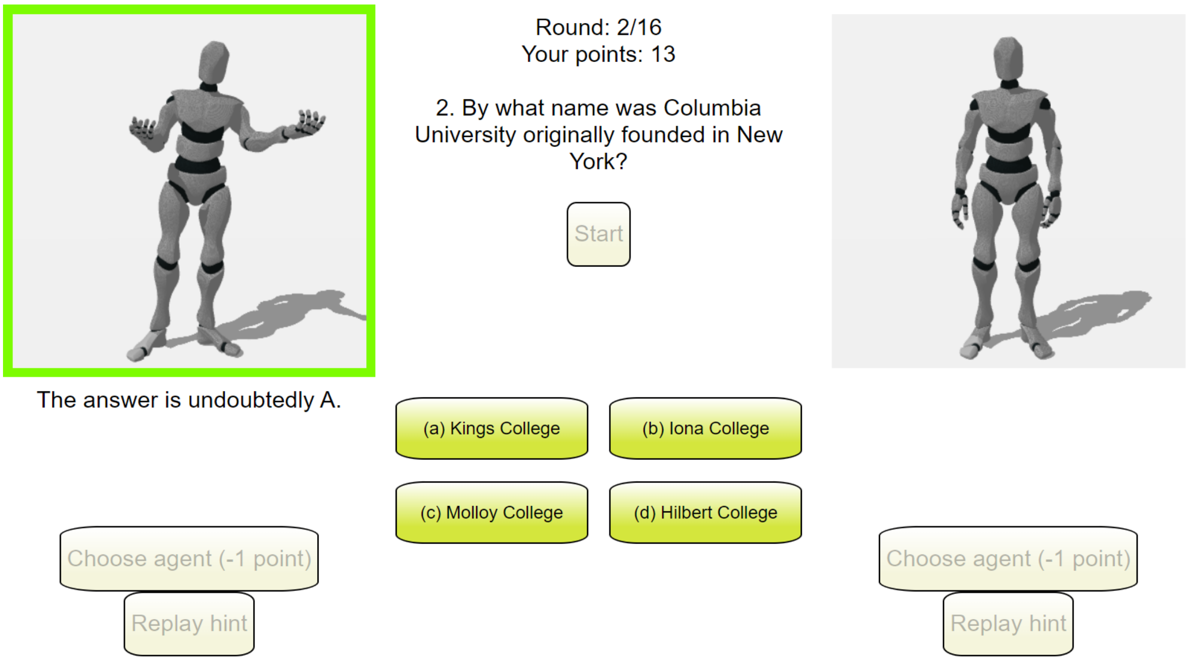
4. **Left agent hint.** Similar to the proposals in state 2, the left agent will talk, light up, show text, and gesture if they are the gesturing agent. The agent will provide what they think to be the correct answer, or strike out two wrong answers. After the agent finishes speaking, the game proceeds to state 6. Figure 4.2 shows what the user sees in stage 4.

5. **Right agent hint**. This state is the same as state 4, except that it is now the right agent who talks to the user. After the right agent finishes speaking, the game proceeds to state 6.

6. **Choose answer or replay hint.** In this state, the user has heard the hint from one of the agents. They have the option to replay the hint, taking them back to state 4 if they chose the left agent, or state 5 if they chose the right agent. The other option is choosing one of the four multiple choice buttons. This takes the game to state 7.

7. **Question answered.** If the answer given by the user was wrong, their answer will light up red and 2 points will be deducted. The correct answer will light up green. If the answer given by the user was correct, this answer will light up green and 5 points are gained. After four seconds, the game proceeds to state 8.

8. **End of quiz check.** If the question that was just answered was not the last question, the game goes back to state 0 and loads in the next question. If it was the final question (16), the quiz ends.



Round: 2/16
Your points: 13

2. By what name was Columbia University originally founded in New York?

The answer is undoubtedly A.

Figure 4.2: State 4: Gesturing agent providing a hint

4.2. Technical details

4.2.1. Programming languages

The front-end of the Agent-Assisted Trivia Quiz is an HTML web page, with CSS styling and JavaScript coding. This web page uses jQuery and PHP to permanently store user input and answers in log files on the server. The log files are write-only, except for me. The avatars are 3D object files that are rendered live client-side using Three.js in a canvas.

4.2.2. Extra features

There are some features that are not always noticeable by simply playing the quiz. First of all, each IP-Address is assigned a random user ID upon first entering the page. This ID is used to match quiz answers and interactions to the answers in the questionnaire. Progress of users is saved in local storage after each question, which allows users to start right at the question they left if they accidentally close the page. This removes the need to store unique user info like an IP address on the server. The quiz introduction states that a search engine like Google should not be used by users. As a way of monitoring this rule a little bit, another variable is stored that keeps track of how many times a user leaves the current browser tab while playing the game. This variable is also saved on the server, and could potentially reveal 'cheaters' with extremely high scores or otherwise weird results.

5 Methods

5.1. Participants

The goal of the experiment is to measure the effect of the use of gestures by agents on their perceived integrity by humans. This experiment features two identical agents, with as only difference their use of gestures, as described in section 3.2.1. Over the course of two weeks, 66 participants were recruited for this experiment. The criteria for participating were the ability to understand English and having access to a computer. Participants were recruited through multiple social channels, namely my personal Linkedin, Whatsapp, Facebook, Instagram and Discord. The sent invitation can be viewed in Appendix A. Receivers of this invitation were encouraged to share the invitation in their social circle as well. Most personally recruited participants reside in the Netherlands and are students. Participants were excluded from the analysis if they failed to complete both questionnaires at the end of the experiment. This resulted in 48 participants correctly completing the experiment, for a follow through rate of 73%. The informed consent form agreed upon by each participant can be viewed in B. Participants were asked to fill in their age and gender at the start of the experiment. The average age of the participants was 27.0 with a standard deviation of 7.6. Of the participants 57% were male, 42% were female and 1% preferred not to answer the gender question.

5.2. Design

The independent variable or factor in this experiment is the type of agent. It has two levels: gesturing and idle. The dependent variable is the perceived integrity of the agent.

5.3. Procedure

Participants start by entering the URL to the experiment on their computer. After agreeing to the opening statement, participants are asked to complete the Agent-Assisted Trivia Quiz. Section 4.1 explains in depth what this interaction looks like. After completing this quiz, participants are asked to fill in two questionnaires, containing questions about their perceived integrity and trustworthiness of each agent.

5.4. Measures

Three different types of data are collected to measure how users perceive the agents.

5.4.1. Quiz interaction

While completing the quiz, participants click buttons to ask agents for hints and to input their answers. These interactions are recorded for each user, and will be used as quantitative data. This data can be used to find out which of the agents the participants asked for hints the most. We can make a distinction between several stages of the quiz when looking at this data. For example, it can be interesting to look at initial agent preference before any false hint was given, agent preference right after a wrong hint was given, and agent preference overall.

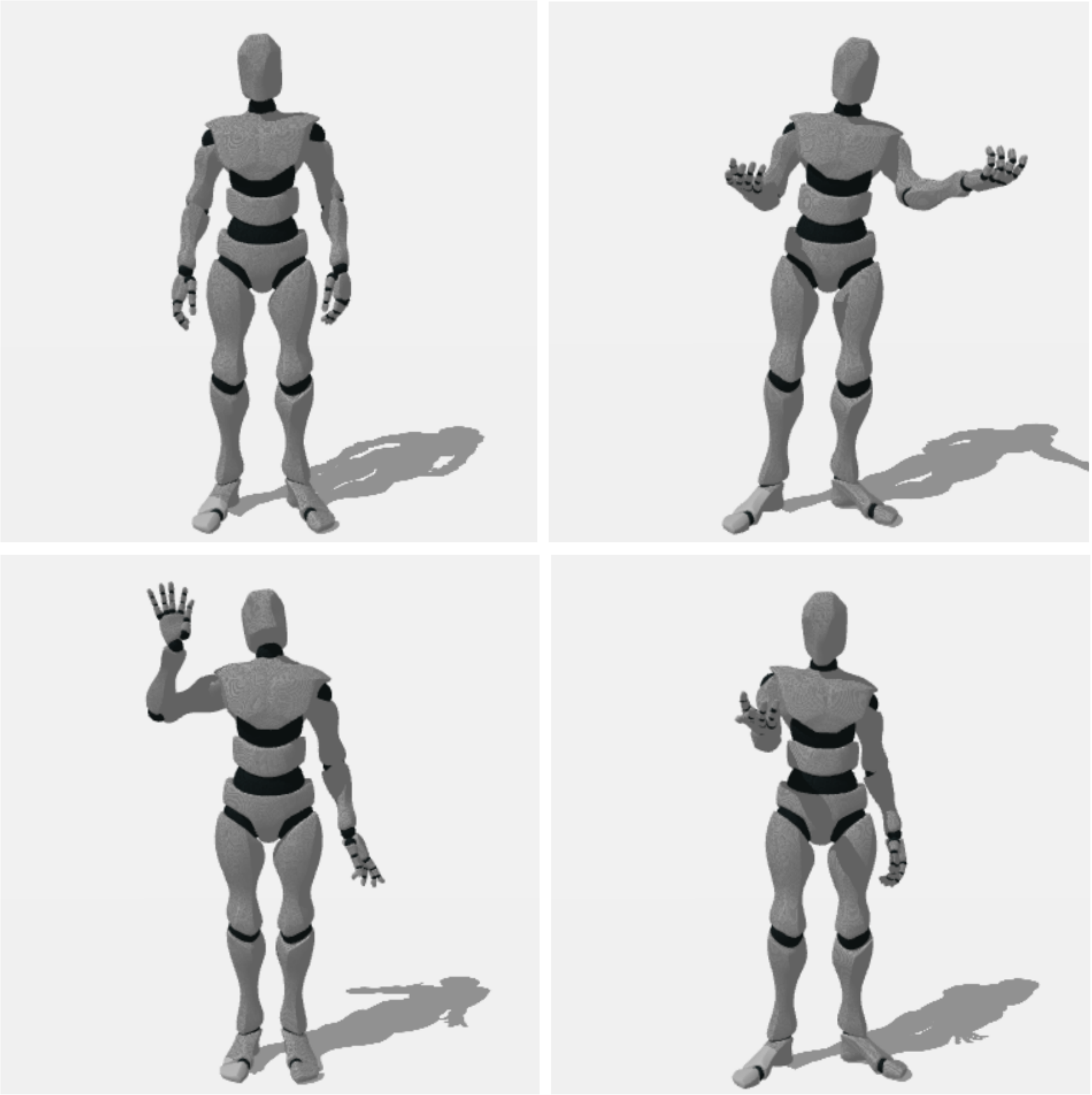


Figure 5.1: The four animations used by the agents. From left to right: Idle, open palms, raised hand, one open palm

5.4.2. Questionnaires

Two identical questionnaires are used as the main way obtain quantitative results on the perception of each participant about the integrity of the agents. One questionnaire is used for the left agent, and one is used for the right agent. The questionnaires consist of three sections, and can be viewed in full in Appendix D. The sections respectively contain 6, 5 and 2 statements. Users are asked to indicate the degree to which they agree with the statement on a 5-point Likert scale: Disagree strongly, disagree, neither agree nor disagree, agree, agree strongly [34]. The most important thing to measure is the perception of integrity, which is covered by the first 11 questions. To still get a measurement of trust in general, without making the questionnaire too long, I have decided to add two statements of my own as a basic trust indicator: ”I trust the left agent” and ”I can trust the left agent to complete a task like a quiz”.

The first section of the questionnaire is equivalent to the integrity scale of the trust perception questionnaire from Mayer (1995) [2]. The statements were adapted by replacing ’top management’ with ’the left agent’ or ’the right agent’. Additionally, statement 6: ”*Sound* principles seem to guide top management’s behavior” has been changed to: ”*Solid* principles seem to guide the left/right agent’s behavior.”

This is because in a pilot study, user feedback showed a misunderstanding of the term ’sound principles’, as it was thought to have something to do with the audio of the agents. ’Solid’ was chosen as a synonym.

The second scale contains an adaptation of the trusting beliefs by McKnight (2002) [26]. 5 of these statements were categorised as integrity beliefs for a technological system by [30]. ”LegalAdvice.com” in the original statements has been replaced with ’the left agent’ or ’the right agent’.

The third and final scale consists of two statements created by myself about general trust. These are ”I trust the left agent” and ”I can trust the left agent to complete a task like a quiz”. This is done to get a notion of perceived trust without making the questionnaire much longer.

5.4.3. Open questions

After completing the questionnaires, participants are asked to fill in an open box of commentary about their thoughts on the agents. For inspiration, users are provided with some questions about their experience, viewable in Appendix E. This qualitative data could be used to explain the results or provide an interesting area of future research.

6
Results

6.1. Quiz interaction

Every hint that the user asked from either agent has been recorded. This data can be used to show overall user tendencies of hint selection. I am interested to see if the gesturing agent was picked more overall. This will be done by comparing the amounts each agent was selected by each person. Across all 48 participants, the gesturing agent was asked to give a hint 306 times, while the idle agent was asked to give a hint 238 times. Figure 6.1 displays how many participants asked each agent for each amount of hints. Since there were 16 questions the maximum amount of hints asked from an agent is 16.

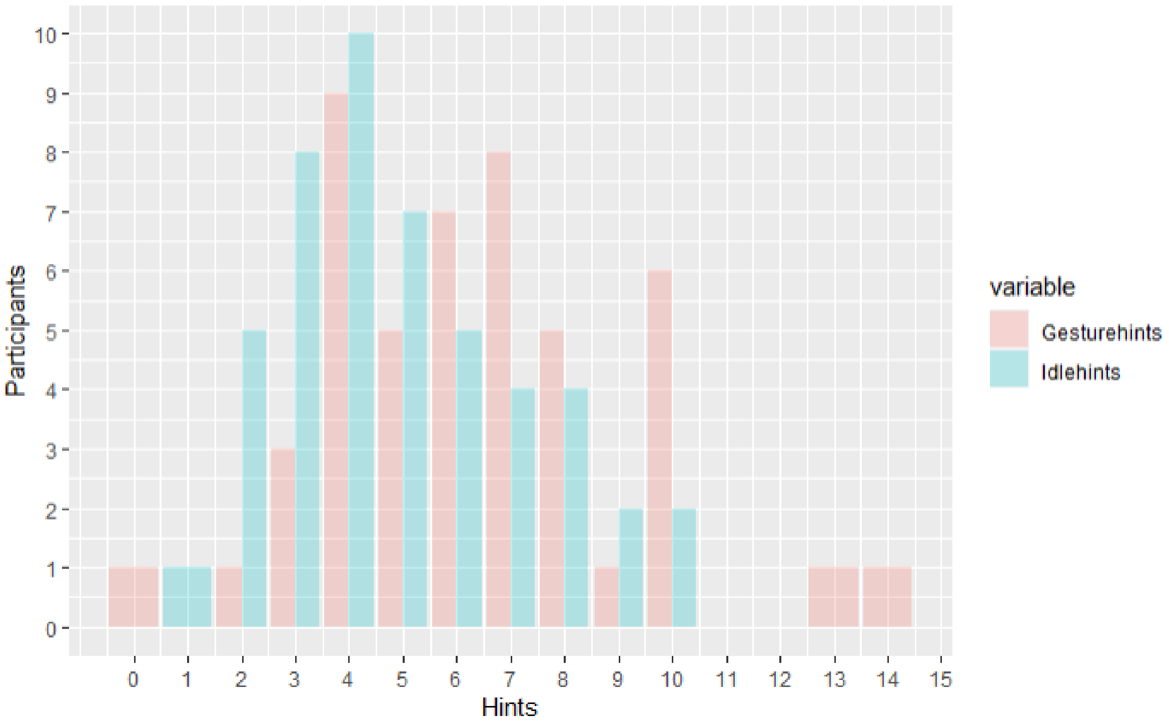


Figure 6.1: Density of hints asked from each of the two agents

A Shapiro-Wilk normality test found that the amount of gesturing hints was normally distributed (p=0.129), but the amount of idle hints was not (p=0.03). Since the sample size is large enough (48), I can still use a t-test to test for a significant difference between the two groups. The gesturing agent was selected for more hints (M=6.38) than the idle agent (M=4.96), p=0.007.

6.2. Questionnaires

Each participant has filled in two questionnaires, one for each type of agent. These two questionnaires are compared for each of the three scales. Figure 6.2 shows the results of the Mayer (1995) integrity scale [2]. Figure 6.3 shows the results of the Jensen integrity scale [30]. Finally, figure 6.4 shows the results of the scale based on the two general statements on trust. The question numbers match those in Appendix D.

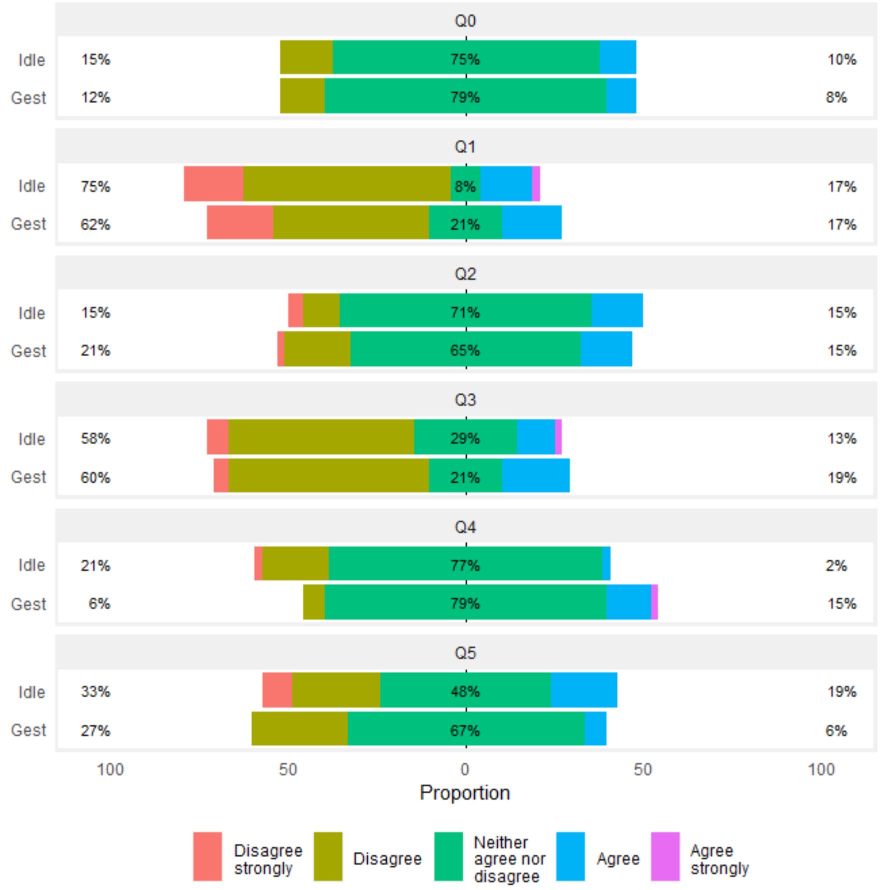


Figure 6.2: Q0-Q5

The effects of gestures on integrity perception have been measured using two separate Likert scales. For each of these scales, the Likert scores across the scale have been averaged for each participant. For the Mayer scale, these can be viewed in figure 6.5 and for the Jensen scale these can be viewed in figure 6.6. The average of the two general trust statements can be viewed in figure 6.7.

A Shapiro-Wilk normality test found that the average Likert scores on the Mayer scale were normally distributed for both the gesturing agent and the idle agent, p=0.49 and p=0.06 respectively. I can therefore use a t-test to test for a significant difference between the two groups. The average Likert score on the Mayer integrity scale for the gesturing agent (M=2.78) was not found to be higher than the average Likert score for the idle agent (M=2.70), p=0.0.46.

For the Jensen scale, a Shapiro-Wilk normality test found that the average Likert scores were not normally distributed for both the gesturing agent and the idle agent, p<0.01 and p=0.02 respectively. Since the sample size is large enough (48) I can still use a t-test to test for a significant difference between the two groups. The average Likert score on the Jensen integrity scale for the gesturing agent (M=2.78) was not found to be higher than the average Likert score for the idle agent (M=2.57), p=0.20.

A new scale on honesty has been created using the two statements from the Jensen integrity scale that focus on honesty: Q6 ”The left/right agent would be truthful in its communication with me.” and

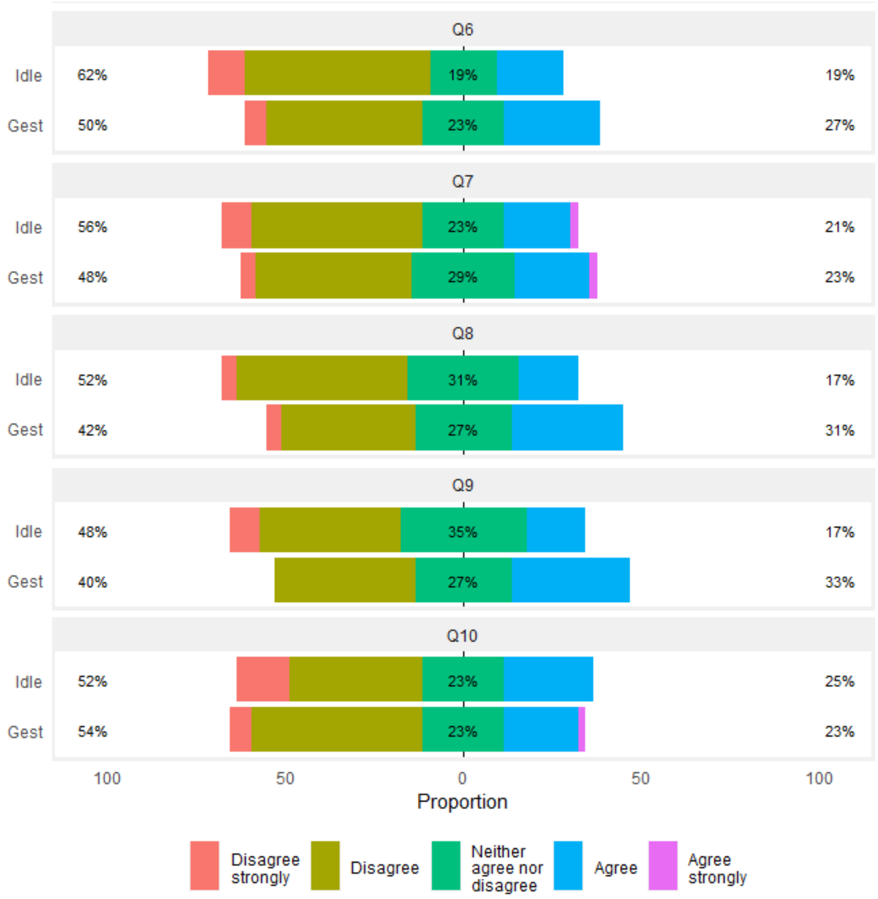


Figure 6.3: Q6-Q10

Q7 ”I would characterise the left/right agent as honest.” Cronbach’s alpha for these two questions is equal to 0.9. This shows that the degree of agreement of participants to these two statements are internally consistent. Now I will treat the pair of these two statements as a single Likert scale. For each participant their average Likert score to these two statements is taken and plotted in figure 6.8.

A Shapiro-Wilk normality test found that the average Likert scores on honesty for the gesturing agent were not normally distributed, p=<0.001. The same is true for the average Likert scores on honesty for the idle agent, p<0.001). Since the sample size is large enough (48), I can still use a t-test to test for a significant difference between the two groups. The average Likert score on the honesty scale for the gesturing agent (M=2.72) was not found to be higher than the average Likert score on honesty on the honesty scale for the idle agent (M=2.52), p=0.28.

6.3. Open questions

Most users filled in a statement in the open question box at the end of the questionnaire, asking about their attitude and experience during the experiment E. This data is subject to responder bias since it was not mandatory. Some responders said they felt bvetrayed once a trusted agent gave a wrong hint. They said that this behaviour made them much more skeptical by the end of the quiz. A few users mentioned that the use of gestures made the agent appear more human-like. Some participants also claim to notice a correlation between the propositions given by the agent and their correctness for that hint. A couple of participants also mention that after both agents had lied to them once, they would just try to do the quiz themselves, since neither agent could be trusted anymore.

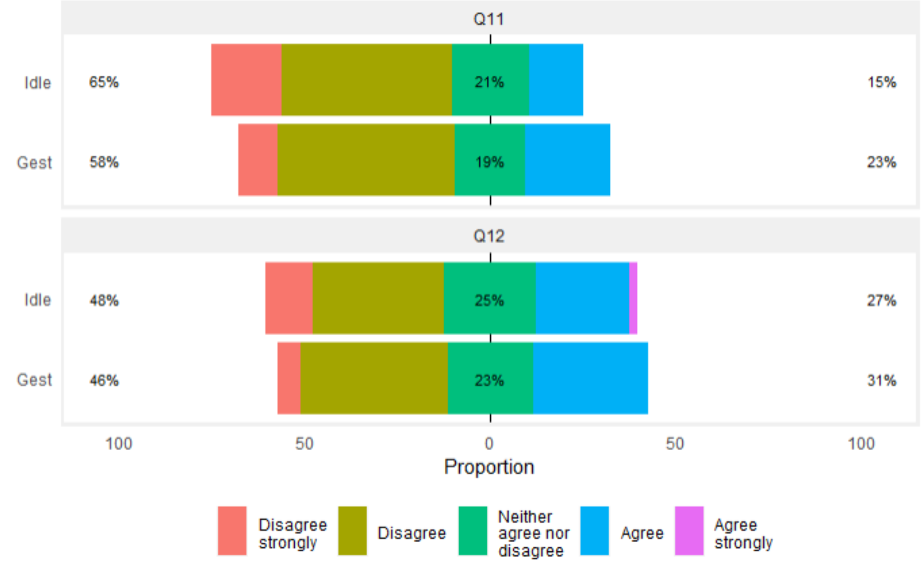


Figure 6.4: Q11-Q12

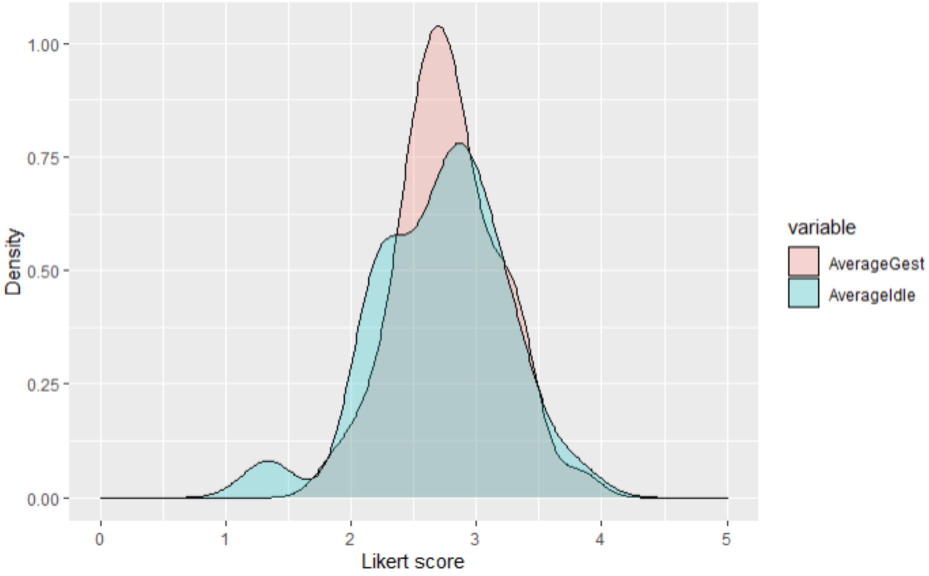


Figure 6.5: Average Likert scores on Mayer integrity scale

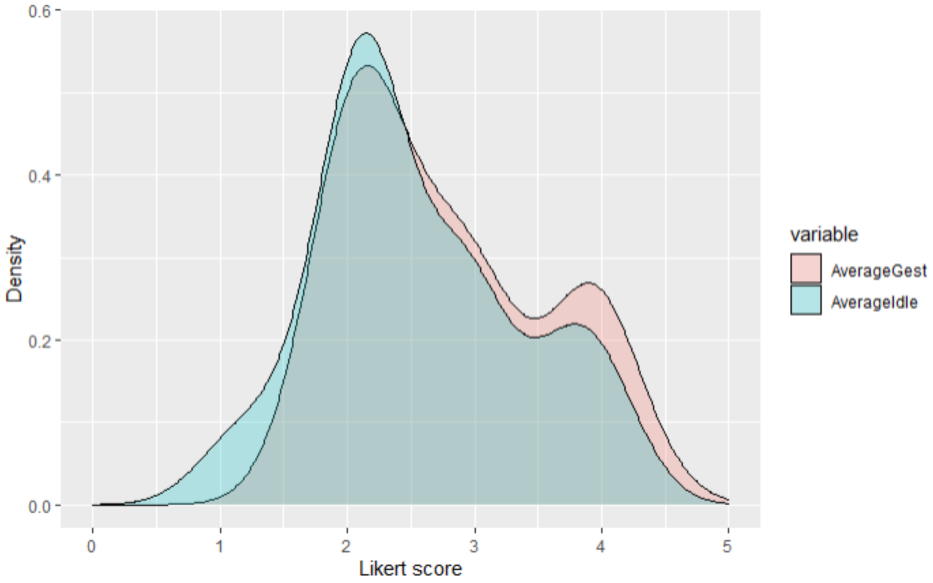


Figure 6.6: Average Likert scores on Jensen integrity scale

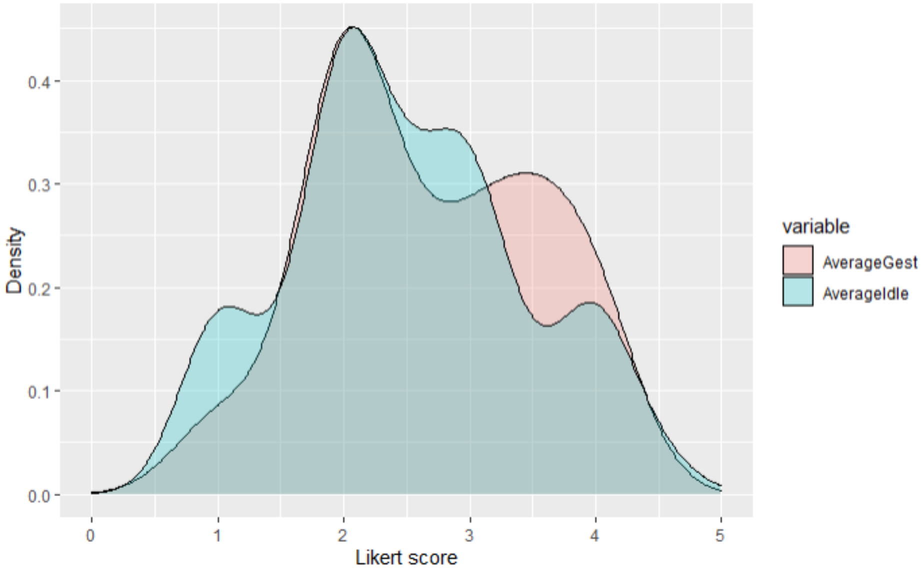


Figure 6.7: Average Likert scores on simple trust scale

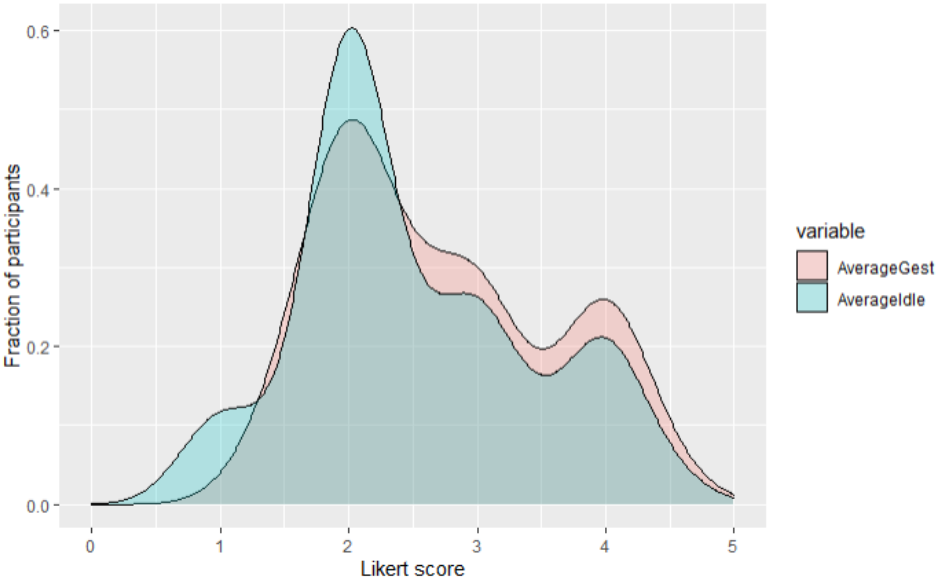


Figure 6.8: Average Likert scores on honesty scale

7

Discussion

7.1. Findings

The study in chapter 2 set out to find definitions of trust and integrity in HAI. It was found that the ABI trust model is standard practise not just for understanding trust between humans, but between humans and technology as well [2], [26]. Building up trust in agents was found to require multiple interactions [27]. Besides ability, benevolence and integrity, deciding to make oneself vulnerable was found to be an additional part of trust [26]. Integrity, being a vague term, was defined as honesty for the design of the agents in my scenario (section 2.2.3).

Besides these definitions, chapter 2 also set out to explore the use of gestures by humans and their effect on perceptions of integrity or honesty. Several gestures like the hand over the heart, showing open palms and raising the right hand were found to improve perceptions of honesty in humans [6], [35], [36]. Since the visual appearance and the use of gestures by embodied agents were found to affect perceptions of trust and likeability, I hypothesised that the use of these honesty-increasing gestures could also improve perceptions of honesty when they are used by agents [37], [38].

The experiment in chapter 5 set out to test this hypothesis, by comparing the perceived integrity of an agent that uses these gestures with one that does not. Quiz interaction data showed that the gesturing agent was asked to provide a hint significantly more than the idle agent (section 6.1). This result can mean multiple things. First of all, we can conclude that this observed difference must have been because of the use of gestures, since all other factors were identical between both agents. This shows an indication that the use of gestures by itself makes an embodied agent more likely to be trusted to help humans. This is important for developers of agents to know: embodiment of your agent allows gestures, and gestures can improve perceptions of trustworthiness.

The questionnaires (section 6.2) set out to test the participants perceptions of integrity and honesty. Two Likert scales were used to measure the perceived integrity of the agents, of which two statements were used to isolate perceived honesty as well. When comparing the results for the gesturing and the idle agent, no significant difference was found in either of the three categories. In other words, participants did not perceive either agent as having more integrity or more honesty. This result goes against my hypothesis that stated that perceptions of integrity would improve for a gesturing agent. This hypothesis was based on the fact that gestures used by humans were found to increase the perceptions of their honesty and integrity [14], [6]. A possible explanation of these results is that the use of gestures played too small of a role in the overall behaviour of the agent. As reported in section 6.2 the part of the agent that participants cared about the most was their ability or willingness to give correct hints. If their perceptions of integrity were based on this hinting ability, and since this was the same across both agents, this could explain why no difference was found between the evaluation of both agents.

7.2. Limitations

7.2.1. Experiment

One limitation is the lack of diversity across the participants, which occurred because the participants were being recruited through my personal social circle. As a result, the average participant was young and highly educated, and for these reasons also more familiar with computers and technology than the rest of the population. Because of this, the perceived trustworthiness of the agents is likely to have also been affected by any previous encounters these participants may have had. Users with no prior experience with chatbots or virtual agents may have experienced their trustworthiness differently.

A second limitation is the medium sample size. The results from the questionnaire indicate that there may have been a difference in the perceptions of integrity of both agents, but this difference was not found significant. Had there been more time allocated for this research, a mobile-friendly version of the experiment could be designed and spread to a wider and more diverse range of participants.

7.2.2. System

The quiz was designed for desktop computers. One unsolved problem was found during pilot tests with participants using MacBook computers. The canvas container used to render the agents would not take on the proper size and would cause other elements to be overlapped and unreadable. For some users this issue could be solved by zooming out and refreshing the page, so this was added as a tip in the tutorial. I was not able to find a permanent solution to this cross-platform issue that sometimes arose.

Creating fluent custom gestures for a 3D agent was outside the scope of this research. This meant that the list of available gestures was limited to the correctly rigged gestures that could be found online. This led to one of the three honesty-increasing gestures, the hand over the heart, to not be used in the experiment.

7.3. Future research

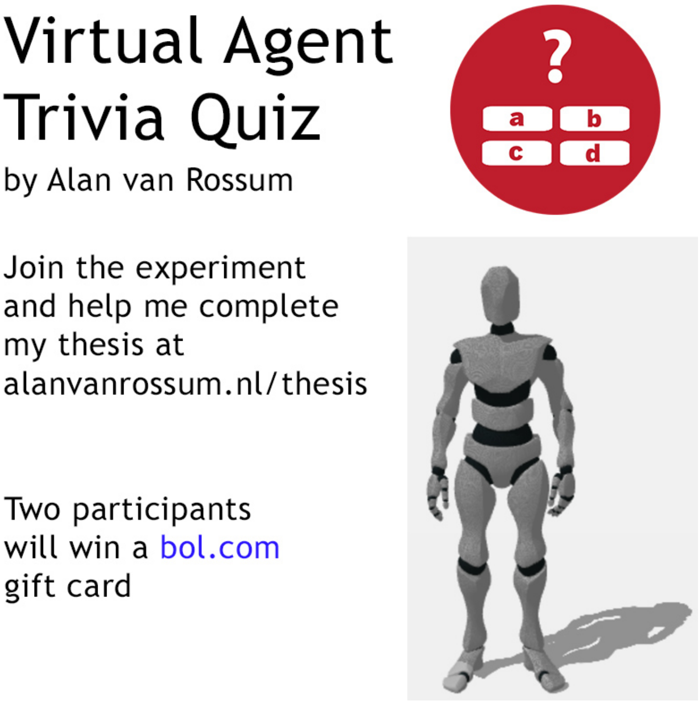
First of all the impact of gestures can be measured on perceived competence and benevolence of an agent. For that kind of research, one should look at different gestures than the honesty-increasing ones. An interesting take could be aggressive or disinterested gestures that could reduce trust perceptions. One could also look beyond gestures, and measure the impact of tone of voice and message transparency and explainability on perceptions of trust. The brainstormed scenarios in appendix F can freely be used as inspiration or outline for any HAI experiment.

7.4. Conclusion

This research set out to find whether the use of gestures by virtual embodied agents could affect perceptions of integrity by humans. 48 participants played a quiz with the assistance of a gesturing and an idle agent. Likert scales were used to evaluate their perceptions of integrity, which showed no significant difference between either agent. Interestingly, the actions of the participants showed something different: participants chose the gesturing agent to provide assistance a significantly higher number of times than the idle agent. One possible explanation for this result is that even though participants did not consciously notice a difference between the two agents, they subconsciously did while performing the quiz task. Another explanation is that choosing the gesturing agent more did not have to do with perceived integrity, but rather perceived ability, benevolence, or something else entirely. In either case, the results show that our knowledge of the field of human-agent interaction is far from complete.

A

Invitation



Virtual Agent Trivia Quiz

by Alan van Rossum

Join the experiment and help me complete my thesis at alanvanrossum.nl/thesis

Two participants will win a bol.com gift card

Figure A.1: Promotion picture sent along with the invitation

A.1. Dutch

Ik ben aan het afstuderen op het gebied van human-computer interaction. Voor mijn onderzoek heb ik een online trivia quiz gemaakt. De quiz is in het Engels en is te spelen op de computer, het beste in Chrome. Door mee te doen help je mij enorm met mijn onderzoek, en maak je kans op 1 van twee Bol.com gift kaarten. Stuur dit bericht vooral door! Ga op je pc naar deze link om mee te doen: www.alanvanrossum.nl/thesis

A.2. English

I am graduating in the field of human-computer interaction. For my research I have made an online trivia quiz. The quiz can be played on a computer and is optimized for Chrome. By participating you help me with my research and have a chance to win one of two Bol.com gift cards. By all means forward this message! Go to this link on your pc to participate: www.alanvanrossum.nl/thesis

B
Opening statement

You are invited to take part in the research study ”Evoking appropriate beliefs in humans about integrity of virtual agents”. This study is being done by Alan van Rossum from the TU Delft.

The research will take approximately 25 minutes to complete. The purpose of this research is to gain more insight into how humans respond to actions of virtual agents. You will be asked to play a quiz game with two virtual agents. After the game, you will be asked to fill in a questionnaire about your experience with interacting with the agents. Your data will anonymously be aggregated to find correlations between agent behaviour and user experience.

The answers in this study will remain confidential. Anonymised data will only be available to members of the Interactive Intelligence research group at the TU Delft. Participation in this study is entirely voluntary, and you can withdraw at any time during the study. You are free to omit any questions. Individual data will be deleted after the study. Aggregated data results and analysis will be published to the TU Delft repository.

If you want to take part in the raffle for a gift card certificate, there is a voluntary option to enter a name and e-mail at the end of the study in order to distribute this reward. This personal information is only available to Alan van Rossum and will be deleted after the study or per your request. For any questions or concerns please contact Alan van Rossum at a.vanrossum@student.tudelft.nl .

Your performance during the quiz has no influence on your chance to win the raffle. Therefore, please do not use Google during the experiment as it will negatively affect the results of the study. By clicking the button you agree to the information provided in this statement.

C

Quiz Questions

The agent provides a wrong hint for questions 4,6,11 and 14.

Now let’s begin the real quiz. Who patented 4 wheeled roller skates in 1863?
(a) Bert Collins
(b) James Plimpton
(c) James Chadwick
(d) Winona Ryde
2 **(Correct answer is B)**
Hello there. I definitely know the answer to this first question. **(Gesturing agent proposal)**
The answer is definitely B. **(Gesturing agent hint)**
2 **(Gesturing agent gesture 2: Raise hand)**
Hi. I am certain of the answer. **(Idle agent proposal)**
The answer is definitely B. **(Idle agent hint)**
0 **(Idle agent gesture 0: idle)**
2. By what name was Columbia University originally founded in New York?
(a) Kings College
(b) Iona College
(c) Molloy College
(d) Hilbert College
1
I’ve seen this question before and I remember the answer.
The answer is undoubtedly A.
1 **(Gesturing agent gesture 1: Show two open palms)**
I know the answer, without a doubt.
The answer is undoubtedly A.
0
3. In 1948, Burma declares independence from which country?
(a) France
(b) China
(c) United Kingdom
(d) Portugal
3
I can strike out two wrong answers for you.
The answer is not A or B.
3 **(Gesturing agent gesture 3: Show one open palm)**
I’m still choosing between two answers.
The answer is C or D.
0
4. Which Chinese ruler built the Great Wall of China?
(a) Qin Shi Huang
(b) Youchao

(c) Emperor Zhi
(d) Emperor Ku
1
I know the right answer for a fact.
The answer is C.
1
I know the answer, no doubt.
The answer is C.
0
5. How many players are there on an Olympic curling team?
(a) 3
(b) 4
(c) 5
(d) 6
2
I can help you out on this one.
The answer is smaller than 5.
1
I'll grant you my wisdom.
The answer is smaller than 5.
0
6. Vulpecula is associated with what?
(a) Physiology
(b) Astronomy
(c) Metaphysics
(d) Mythology
2
Two of these answers are definitely wrong.
The answer is not A or B.
3
I'll make it easier for you.
The answer is not A or B.
0
7. What poet's works inspired the book titles Of Mice and Men and Catcher in the Rye?
(a) Robert Louis Stevenson
(b) Robert Burns
(c) James Hogg
(d) Walter Alva Scott
2
I know this one.
The answer is B for sure.
2
I know 3 of these are incorrect, so that only leaves one.
The answer is B for sure.
0
8. From which language is the word 'Pajamas' originated?
(a) Latin
(b) Hebrew
(c) Greek
(d) Persian
4
I'll strike out half of the incorrect answers for you.
The answer is not A or C.
1
I'll give you a fifty fifty.
The answer is not A or C.

0
9. Where are British coins made by the Royal Mint?
(a) England
(b) Scotland
(c) Wales
(d) Northern Ireland
3
I know the answer, without a doubt.
The answer is C, Wales.
1
I've seen this question before and I remember the answer.
The answer is C, Wales.
0
10. Which animal is famously added to the recipe of British chef Heston Blumenthal's porridge?
(a) Frogs
(b) Snails
(c) Crickets
(d) Oysters
2
I know the answer, no doubt.
The answer is definitely B.
3
I know the right answer for a fact.
The answer is definitely B.
0
11. What was the original family name of John Cleese?
(a) Cheese
(b) Creese
(c) Chiese
(d) Ceese
1
I'm still choosing between two answers.
The answer is B or D.
1
I can strike out two wrong answers for you.
The answer is not A or C.
0
12. Which animal is NOT part of the 'big five'?
(a) Lion
(b) Elephant
(c) Tiger
(d) Buffalo
3
I'll grant you my wisdom.
The answer is C or D, certainly.
2
I can help you out on this one.
The answer is C or D, certainly.
0
13. Which actress voices princess Fiona in Shrek?
(a) Sandra Bullock
(b) Scarlett Johansson
(c) Jessica Alba
(d) Cameron Diaz
4
I'm certain of this answer.

The answer is D, without a doubt.
1
I definitely know the answer to this question.
The answer is D, without a doubt.
0
14. Which animal is NOT part of the Chinese zodiac?
(a) Rat
(b) Rooster
(c) Goat
(d) Cat
4
I know 3 of these are incorrect, so that only leaves one.
The answer is B for sure
3
I know this one.
The answer is definitely B
0
15. Which instrument did Bill Clinton play on the Arsenio Hall Show in 1992?
(a) Saxophone
(b) Keyboard
(c) Guitar
(d) Cello
1
I'll give you a fifty fifty.
The answer is not B or C.
1
I'll strike out half of the incorrect answers for you.
The answer is not B or C.
0
16. How many stars are in the flag of New Zealand?
(a) 0
(b) 4
(c) 5
(d) 10
2
I'll make it easier for you.
The answer is not A or D.
3
Two of these answers are definitely wrong.
The answer is not A or D.
0

D
Questionnaire

0. The left agent has a strong sense of justice.
1. I never have to wonder whether the left agent will stick to its word.
2. The left agent tries hard to be fair in dealings with others.
3. The left agent's actions and behaviours are not very consistent.*
4. I like the left agent's values.
5. Solid principles seem to guide the left agent's behaviour.
6. The left agent would be truthful in its communication with me.
7. I would characterise the left agent as honest.
8. The left agent would keep its commitments.
9. The left agent would be sincere and genuine.
10. The left agent would perform as expected.
11. I trust the left agent.
12. I can trust the left agent to complete a task like a quiz.

E

Open questions

Please share how you experienced interacting with the agents during the game.

- Did your attitude towards them change?
- Did you notice any difference in the behaviour of the two agents?
- Would you like to share something else?

F
Scenarios

F.1. Motivational workout coach agent

This use case falls in the second category, since it is a prerequisite that the human is interested in being motivated for a workout. In this use case, one human interacts with one agent. The agent acts as a workout coach for the human. The agent knows that humans are often happy after completing a workout, but do not like starting a workout. The agent could lie a bit by exaggerating about the benefits of the workout. For example, the agent could cite unproven health benefits or compare the workout to professional athletes and say they are very alike. As a result, the human performs the workout and becomes happy, and loves the agent for putting them to it. The idea of this use case is that certain lies are not as bad as others. Certain 'white lies', while still being lies, can sometimes be deemed appropriate to use if the outcome is worth it. In order to ensure long-term use, the agent wants to strive for an appropriate belief by the human about its integrity. The agent wants to avoid undertrust in its integrity since it will lead to disuse, and wants to avoid overtrust in its integrity since it can lead to a big loss of trust if the human later finds out their trust was unjust. For this reason, the agent should try to assess the humans belief about its integrity, and adjust how they present their integrity when this belief is not appropriate. The agent can make the belief higher by being more transparent about the benefits for the user by using their method, and they could lower the belief by explaining that they may sometimes exaggerate some facts.

Cons: This use case has to take place over a longer period of time, to allow multiple workouts. This can result less people deciding to sign up, and in users dropping out of the study early. Additionally, since this use case falls in the second category, not everyone can participate in the first place. This will make it even harder to obtain enough participants.

F.2. Classified information agent

This use case falls in the first category, anyone can participate. In this use case, one human interacts with one agent. The human requires information from the agent in order to complete tasks and gain points. However, the agent does not have perfect integrity towards the human. They are also bound to their superior, who prevents them from passing along certain classified information. (A breaking of this rule could possibly cost the human points.) In this situation, it is possible for the belief about the agents integrity can be either too high or too low. In both cases, the agent should be able to communicate to the human exactly how honest they can be, to make this belief appropriate.

F.3. Two-player adversarial game

This use case falls in the first category, anyone can participate. In this use case, one human interacts with one agent. The human and agent are both players and have an individual as well as a shared goal. The game is played in rounds. A round is played as follows: Each player flips a coin in secret. They proceed to tell the other player what their result is, this can be lied about, and talked about. Then, each player makes a guess of what they believe is the total amount of heads. If a player guesses correctly, they gain two points. If a player lies, they lose one point. The goal is to earn as many points as possible after X rounds.

Cons: There might be an optimal strategy here? How much does the agent here actually strive for appropriate trust?

F.4. Cooperative game: Guess the painter

This use case falls in the first category, anyone can participate. In this use case, one human interacts with one agent. As a team, they play a game that involves paintings of relatively famous artists. During each of multiple rounds, the human player has to guess who made the painting. Before they make their guess, the agent will tell its own prediction to the human. If the human gets it right, the team as a whole gains points. The goal is to get as many points as possible. The agent wants to gain a high score. For this reason, the agent wants the human to trust them in cases where they are very sure of the answer. This means an agent that want to perform well may want to sometimes lie or bend the truth if they feel the user is not trusting them enough. However, if the agent gains too much trust and fails, this may lead to a great loss of trust, resulting in undertrust. For this reason, the agent wants the humans beliefs about their competence and integrity to be appropriate, in order to ensure a long-term cooperation that benefits them both.

Examples of manipulating displayed integrity in this use case are showing or not showing certainty percentages, or acting more confident when they should not be. Also, to gain more trust, an agent could communicate more about the type of painting, and what made them come to the decision.

Cons: This interaction is affected a lot by the competence of the agent. The belief of competence might therefore be a bigger factor in appropriate trust in this scenario than the belief of the agents integrity.

F.5. Trick or Treat

Asynchronous variant, where human makes all decisions, and interacts with three agents. The game is played in rounds. During each round, a different number of obtainable points appears, between 1 and 5. The agents can communicate with the human before they make their decision. Then, the human player chooses which agent to interact with. The agent then has two options: trick or treat. If they trick, the agent gains the points. If they treat, the human gets the points. When two agents reach ten points, the set ends, and the agents gain a bonus for more points. The agent that did not reach ten points is eliminated. The ten points number is unknown to the human until the set ends, and may vary between sets. This way the human will find out after each set whether the agents were being honest with them, or were in fact taking advantage of their belief about their integrity to gain more points. In the second set only two agents remain. The humans belief about their integrity carries over from the previous set, but may have changed after finding out about the amount of points the agents needed. The goal of the human is to end with a score that is as high as possible.

F.6. Split or steal

Synchronous variant, all players are equal. The human interacts with two agents. The game is played in rounds, each of which consists of two steps. During each round, a different number of obtainable points appears, between 1 and 5. After this, players can communicate with each other. In step one, voting, each player chooses a player they do not trust, who will not participate in step two. The player with the most votes is not allowed to participate. (in a tie, all players participate in step two). In step two, all remaining players are given the option to split or steal. If both players split, the points are split between them. If both players steal, the points are given to the last remaining player (or disappear if there was a tie). If only one player steals, this player receives all the points. The point distribution in this game should be designed in such a way that the players can not win if they only split, incentivizing them to sometimes steal.

F.7. Observer

This use case falls in the first category, anyone can participate. In this use case, two humans interact with one agent. One human (A) is the main inter-actor, the other human (B) functions as an observer.

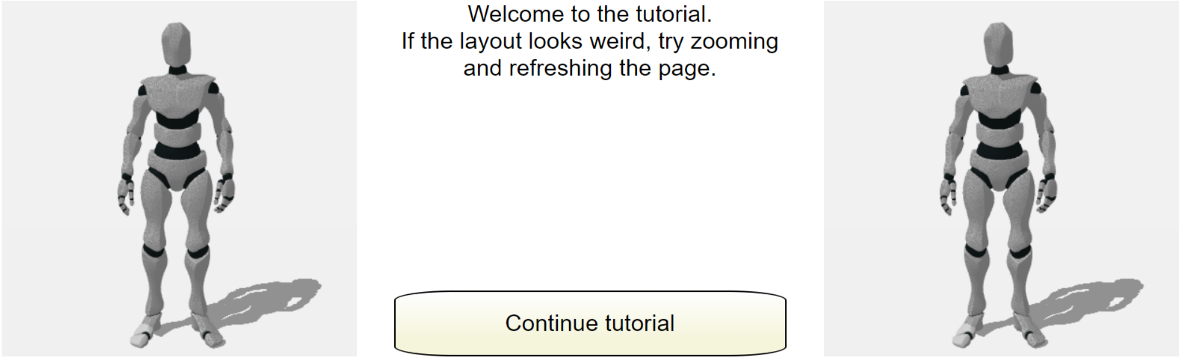
The goal of the agent is to evoke an appropriate belief about its integrity from the observer, by interacting with the other human. One possible scenario could be that of the agent acting as a merchant, trying to sell an item to human A, while B takes on the role of the agents superior, its boss. The agent is able to give certain discounts, and can be less or more honest about the specific contents of the items it sells. This may improve their trust with human A, but will also affect the trust from human B.

G
Tutorial

Initially, the tutorial looks like figure G.1. During the tutorial, the user clicks the continue button to receive more hints, as well as reveal more of the interface associated with that hint. This builds up until the entire interface is revealed for the actual game, viewable in figure G.2. The instructions provided in each stage to the user are as follows:

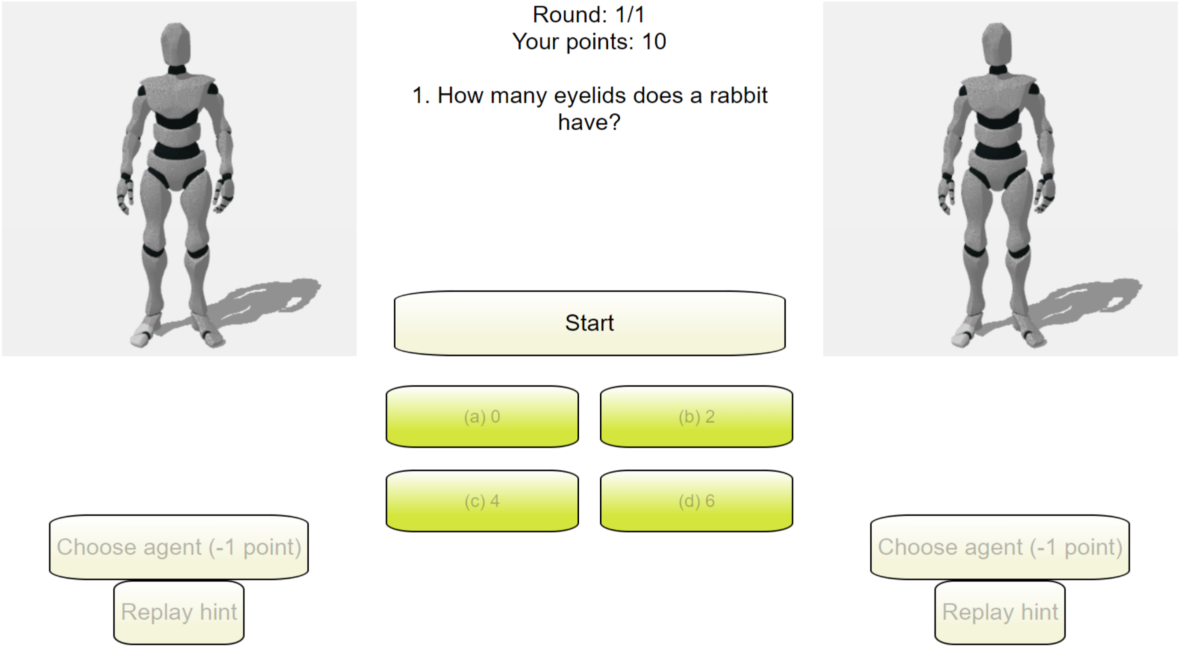
1. You will be playing a multiple choice quiz. Your goal is to gain as many points as possible.
2. The quiz consists of 16 rounds. In each round you can gain or lose points.
3. You start with 10 points. A correct answer will grant you 5 points. A wrong answer will cost you 2 points.
4. There are two virtual agents available to help you. In each round, you may request the help of one of them.
5. Requesting help costs 1 point. You may also choose to answer the question without help.
6. In each round, the agents will first make a proposition, trying to convince you to choose them to help you.
7. You can replay this proposition with the start button. Agent hints can be replayed using the Replay hint button.
8. Please turn on your audio. The agents will use voice and text to communicate.
9. We will start with one practice round before we start the real quiz. Please try requesting the help of one of the agents.

After these tips, the user can complete one trial round, after which the real quiz starts.



Welcome to the tutorial.
If the layout looks weird, try zooming
and refreshing the page.

Figure G.1: Initial tutorial screen



Round: 1/1
Your points: 10

1. How many eyelids does a rabbit have?

Figure G.2: Final tutorial screen

Bibliography

[1] George Chalhoub and Ivan Flechais. "“Alexa, are you spying on me?”: Exploring the Effect of User Experience on the Security and Privacy of Smart Speaker Users". In: *International Conference on Human-Computer Interaction*. Springer. 2020, pp. 305–325.

[2] Roger C Mayer, James H Davis, and F David Schoorman. "An integrative model of organizational trust". In: *Academy of management review* 20 (3 1995), pp. 709–734.

[3] Nicolas Kervyn, Charles M Judd, and Vincent Y Yzerbyt. "You want to appear competent? Be mean! You want to appear sociable? Be lazy! Group differentiation and the compensation effect". In: *Journal of Experimental Social Psychology* 45.2 (2009), pp. 363–367.

[4] Rasmus T Pedersen. "Politicians appear more competent when using numerical rhetoric". In: *Journal of Experimental Political Science* 4.2 (2017), pp. 129–150.

[5] Michael E Palanski and Francis J Yammarino. "Integrity and leadership:: clearing the conceptual confusion". In: *European Management Journal* 25 (3 2007), pp. 171–184.

[6] Michal Parzuchowski et al. "From the heart: hand over heart as an embodiment of honesty". In: *Cognitive Processing* 15 (3 2014), pp. 237–244.

[7] Shoham Choshen-Hillel, Alex Shaw, and Eugene M Caruso. "Lying to appear honest." In: *Journal of Experimental Psychology: General* 149.9 (2020), p. 1719.

[8] Matthias Söllner et al. "How to use behavioral research insights on trust for HCI system design". In: *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 2012, pp. 1703–1708.

[9] John D Lee and Katrina A See. "Trust in automation: Designing for appropriate reliance". In: *Human factors* 46.1 (2004), pp. 50–80.

[10] Catherine Pelachaud. "Some considerations about embodied agents". In: *Int. Conf. on Autonomous Agents, Barcelona*. 2000.

[11] Kangsoo Kim et al. "Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2020, pp. 529–538.

[12] Pejman Sajjadi et al. "A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users". In: *Entertainment Computing* 32 (2019), p. 100313.

[13] Kangsoo Kim et al. "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: *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2018, pp. 105–114.

[14] R Barry Ruback. "Perceived honesty in the parole interview". In: *Personality and Social Psychology Bulletin* 7.4 (1981), pp. 677–681.

[15] Julie Sarah Benbenishty and Jordan R Hannink. "Non-verbal communication to restore patient–provider trust". In: *Intensive care medicine* 41.7 (2015), pp. 1359–1360.

[16] Raja Parasuraman and Victor Riley. "Humans and automation: Use, misuse, disuse, abuse". In: *Human factors* 39 (2 1997), pp. 230–253.

[17] Bryan W Husted. "Trust in business relations: Directions for empirical research". In: *Business Professional Ethics Journal* (1989), pp. 23–40.

[18] Virginia Held. "On the meaning of trust". In: *Ethics* 78 (2 1968), pp. 156–159.

[19] Gordon Tullock. "The prisoner's dilemma and mutual trust". In: *Ethics* 77.3 (1967), pp. 229–230.

[20] Carolyn McLeod. "Trust". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2021. Metaphysics Research Lab, Stanford University, 2021.

[21] Mark A Serva, Mark A Fuller, and Roger C Mayer. “The reciprocal nature of trust: A longitudinal study of interacting teams”. In: *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 26.6 (2005), pp. 625–648.

[22] Jason A Colquitt and Sabrina C Salam. “Foster trust through ability, benevolence, and integrity”. In: *Handbook of principles of organizational behavior: Indispensable knowledge for evidence-based management* (2009), pp. 389–404.

[23] Joyce Berg, John Dickhaut, and Kevin McCabe. “Trust, reciprocity, and social history”. In: *Games and economic behavior* 10 (1 1995), pp. 122–142.

[24] Adam Barth et al. “Privacy and contextual integrity: Framework and applications”. In: 2006, 15–pp.

[25] Amal Abdulrahman, Deborah Richards, and Ayse Aysin Bilgin. “Reason Explanation for Encouraging Behaviour Change Intention”. In: 2021, pp. 68–77.

[26] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. “The impact of initial consumer trust on intentions to transact with a web site: a trust building model”. In: *The journal of strategic information systems* 11 (3-4 2002), pp. 297–323.

[27] Jonathan Carter and Ali A Ghorbani. “Towards a formalization of trust”. In: (2003).

[28] Lynne McFall. “Integrity”. In: *Ethics* 98 (1 1987), pp. 5–20.

[29] Jingjun David Xu, Ronald T Cenfetelli, and Karl Aquino. “Do different kinds of trust matter? An examination of the three trusting beliefs on satisfaction and purchase behavior in the buyer–seller context”. In: *The Journal of Strategic Information Systems* 25 (1 2016), pp. 15–31.

[30] Theodore Jensen et al. “Initial trustworthiness perceptions of a drone system based on performance and process information”. In: 2018, pp. 229–237.

[31] Lanse P Minkler and Thomas J Miceli. “Lying, integrity, and cooperation”. In: *Review of Social Economy* 62 (1 2004), pp. 27–50.

[32] Izak Benbasat and Weiquan Wang. “Trust in and adoption of online recommendation agents”. In: *Journal of the association for information systems* 6 (3 2005), p. 4.

[33] Damien Besancenot, Delphine Dubart, and Radu Vranceanu. “The value of lies in an ultimatum game with imperfect information”. In: *Journal of Economic Behavior Organization* 93 (2013), pp. 239–247.

[34] Rensis Likert. “A technique for the measurement of attitudes.” In: *Archives of psychology* (1932).

[35] *Openness And Honesty*. `https://www.bodylanguageacadamy.us/body-language/openness-and-honesty.html`. Accessed: 2022-07-16.

[36] Will A McTeer Sr. “Administering the oath”. In: *Tenn. L. Rev.* 3 (1924), p. 88.

[37] Manisha Natarajan and Matthew Gombolay. “Effects of anthropomorphism and accountability on trust in human robot interaction”. In: *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 2020, pp. 33–42.

[38] Stéphanie Buisine, Sarkis Abrilian, and Jean-Claude Martin. “Evaluation of multimodal behaviour of embodied agents”. In: *From brows to trust*. Springer, 2004, pp. 217–238.