



The Emotions and Advice in Virtual Assistants: A Dual Study on Emotion Validation and Agent Suggestions in a Gaming Scenario

Raul Paradedá   [Universidade do Estado do Rio Grande do Norte | raul.paradedá@uern.br]

Álison Alves  [Instituto Federal do Rio Grande do Norte | alisoncc.oliveira@gmail.com]

Daniel Torres  [Universidade do Estado do Rio Grande do Norte | danieltorres@alu.uern.br]

Althierfson Lima  [Universidade do Estado do Rio Grande do Norte | althierfsonlima@alu.uern.br]

 Department of Computer Science, State University of Rio Grande do Norte, Av. Dr. João Medeiros Filho, S/N, Potengi (Zona Norte) Natal, RN, Brazil, 59.120-555, Brazil.

Received: 30 September 2023 • Accepted: 19 December 2023 • Published: 01 January 2024

Abstract In an era where virtual assistants play an increasingly prominent role in our daily lives, this study explores the implications of their advice. We investigate the interplay between trust and virtual agents' emotional expressions, delving into a critical aspect of human-technology interaction. Conducted through a comprehensive study comprising two interconnected phases, our research examines the dynamics between virtual agents and human decision-making. The first phase involves developing and validating a virtual robotic agent capable of conveying a spectrum of emotions. Through this, gender-based differences in emotional cue perception are disclosed, shedding light on how men and women interpret these cues differently. The second phase employs an interactive memory game, where the virtual agent operates in varied emotional states. Participants' trust levels and perceptions are meticulously evaluated in different scenarios, ranging from accurate to erroneous agent cues. Our findings elucidate the impact of the agent's emotional expressions on participants' perceptions, illustrating how trust is intricately influenced by both the task at hand and the agent's behavior. This research contributes to understanding the relationship between virtual assistants and human decision-making, emphasizing the necessity of designing more engaging and interactive virtual agents. These insights prepare future research for crafting more effective virtual assistants, fostering increased user trust and engagement.

Keywords: Virtual Assistant; Emotion Validation; Gaming Scenario; Human Trust; Agent Suggestions.

1 Introduction

In today's digital landscape, technology has incorporated itself into the fabric of our daily lives, serving many purposes, from communication (Guzman and Lewis [2020]) to education (Timotheou *et al.* [2023]) and even providing physical assistance (Cicirelli *et al.* [2021]). Within this digital landscape, personal assistance technology has gained prominence as a key player, extending a helping hand to users in various tasks, be it navigating through digital maps or controlling smart home devices through voice commands. Virtual assistants like ALEXA, SIRI, Cortana, and Google Assistant have become household names, yet they are not without their imperfections, occasionally offering incorrect advice or failing to meet user expectations. Such imperfections can have a profound impact on an individual's trust in these virtual assistants (Correia *et al.* [2018]; Lewis *et al.* [2018]).

Although researchers have made significant progress in addressing the faulty behaviors exhibited by virtual assistants (Cuadra *et al.* [2021]), the extent to which such failures or incorrect advice impact an individual's trust and human-agent perception remains a subject of inquiry. This analysis aims to bridge the gap in our understanding, exploring how trust and perceptions of virtual agents are influenced by factors such as anthropomorphism, animacy, likeability, and perceived intelligence. The importance of this investigation extends beyond the domain of our immediate research interests, as it resonates with broader concerns surrounding the effective-

ness of human-technology interactions.

Notably, studies have highlighted the key role of trust in shaping the relationship between individuals and agents. Hancock *et al.* [2011] emphasize the significance of an agent's performance and its attributes, including transparency, in shaping trust dynamics. Moreover, the impact of erroneous behaviors on how humans perceive agents has been a subject of previous research (Ragni *et al.* [2016]; Salem *et al.* [2013]). These findings underscore the relevance of our investigation and the potential ramifications it may hold for the design and implementation of more effective and dependable virtual assistants.

In light of these considerations, our research endeavors to investigate the effects of virtual assistant advice on human trust and perception within the context of digital casual memory games. To achieve this, we conducted two distinct yet interconnected studies. First, we developed a virtual agent named Roboldo, and the initial study focused on validating the emotions expressed by this virtual agent. Subsequently, the second study assessed the impact of the game scenario we developed, where the virtual agent provided both correct and incorrect clues to guide players in card selection.

In this way, to verify if the virtual assistant advice impacts human trust, we developed a virtual agent named Roboldo before developing the game. So, this paper describes the conduction of a research study divided into two phases; the first validates the emotions performed by our virtual agent, and the second measures the impact of our developed game sce-

nario.

In our initial research phase, we employed six animations based on Ekman [1971] study to craft emotive expressions for our virtual agent, and participants were invited to identify the perceived emotions. Additionally, we examined whether the agent's emotions were perceived differently based on the participants' gender. This decision was motivated by previous studies (Hashemian *et al.* [2019]) suggesting that gender can influence the interpretation and recognition of emotional expressions.

After this initial phase, we developed the memory game, where the virtual agent interacted with the participant. This virtual agent, with its simulated emotions (sadness, anger, and happiness), actively sought to persuade participants to select specific cards. To this interaction, we measured the participant's trust levels (Benbasat and Wang [2005]), assessed human-agent perception (Bartneck *et al.* [2009]), and investigated whether the agent's role as an assistant influenced the overall game experience (Poels *et al.* [2007]).

In our research, containing two interconnected phases, we formulated a comprehensive set of hypotheses to address various aspects of human-agent interaction.

- **H1:** Participants correctly perceive the emotions simulated by the agent.
- **H2:** Men and women perceive the emotions simulated by the agent differently.
- **H3:** A person's level of trust towards an agent is affected if the agent's suggestions are incorrect.
- **H4:** False information the agent provides can affect a person's perception of this agent.
- **H5:** People with a level of trust that was not negatively affected by the agent's false suggestions follow the tips provided by the agent more.
- **H6:** The presence of an agent within the game positively affects the player's experience if it provides accurate advice.

In this research, we conduct a study with two interconnected phases. The initial phase serves a dual purpose: it provides comprehensive insights into crafting a virtual agent while concurrently validating the emotions it exhibits. This foundational investigation lays the groundwork for our subsequent phase. In the second phase, participants engage in a digital casual memory game. Leveraging this gaming scenario, we explore factors previously mentioned, such as virtual agent persuasion capability, participant-agent trust levels and perceptions, and player experiences. This research aims to enhance our understanding of how virtual assistants impact human decision-making and reliance, particularly in situations susceptible to guidance errors. Furthermore, this research underscores the critical need for more dependable and trustworthy virtual assistants that align better with user expectations and needs, offering potential implications that extend to domains such as human-robot interaction, human-computer interaction, and cognitive psychology. Specifically, our findings have the potential to inform and enhance the design and functionality of virtual assistants in these fields, facilitating more effective human-technology interactions and furthering our understanding of the intersections between technology and psychology.

The following section (Section 2) reviews relevant literature on trust, human perception of agents, and the impact of erroneous behaviors. Subsequently (Section 3), we elucidate the methodology used for the study and your two phases. As we proceed, we provide insight into the measurement instruments deployed to gauge participants' trust levels, human-agent perception, and the influence of the agent's conduct on the gaming experience (Section 4). Then, we present the results obtained from both phases (Section 5). Next, we discuss the outcomes of both phases (Section 6). Finally, we synthesize our findings and contemplate their implications in enhancing the reliability and engagement of virtual assistants (Section 7).

2 Background and Related Work

In this work, we have developed an agent that leverages a set of bodily movements and behaviors, commonly referred to as emotions (Darwin and Prodger [1998]), as persuasive tools in interactions with players. These emotional expressions must be meticulously designed, precisely defined, and readily interpretable to ensure that individuals can grasp the conveyed sentiments effectively (Johnston and Thomas [1981]). Additionally, the character must incorporate elements, such as body movements or environmental cues, that enable the audience to perceive the conveyed emotions. These elements amplify the character's ability to communicate emotions to the individuals watching or interacting with it. Our research simulates emotions through facial expressions and bodily movements, encompassing emotions like sadness, anger, happiness, surprise, contemplation, shame, and joy.

In this research, we also delve into the field of psychology, focusing on the concept of persuasion and its causes and consequences. Persuasion techniques encompass a broad spectrum, ranging from subtle approaches like verbal communication, gaze, pointing behaviors, or the display of pleasant images to more overt tactics, including physical coercion, threats, or the presentation of unpleasant images (Gass and Seiter [2018]). Importantly, it's well-established that not all individuals respond uniformly to the same persuasion technique (Moyer-Gusé [2008]). In our work, we programmed the agent to persuade participants by blending facial expressions and emotions. For example, if a participant did not follow the agent's advice, the agent would execute a sad bodily movement and facial expression.

In a related study, Hashemian *et al.* [2019] examined the influence and trust in an agent by programming it to employ expressions simulating emotions and sharing a sad story. Findings revealed that participants' trust in the agent increased when the interaction commenced with "small talk", and the agent's display of joy or sadness influenced participants' trust. The study also observed gender-based variations in participants' behavior when exposed to the agent's emotions. Therefore, our work seeks to investigate the effects of simulating emotions in a virtual agent within a casual game scenario, specifically focusing on trust in this context.

In Chowanda *et al.* [2016], researchers explored the player's experience in a game where Non-Player Characters (NPCs) could express and perceive emotions. Question-

naires were employed to gauge the impact of interactions with these agents on player engagement and immersion in the game. The study revealed that participants reported heightened levels of engagement and immersion when interacting with agents that expressed emotions. Furthermore, players could discern specific personality and emotional traits in these agents. Based on these findings, the authors concluded that these emotionally expressive agents enhance the user experience.

In Yang *et al.* [2017], the authors aimed to investigate how individuals perceive emotions in a virtual agent named Zara. The study entailed the development of an agent capable of perceiving certain emotions of users and simulating emotions through animations. Participants were divided into two groups: one interacted with an agent simulating emotions, while the other group interacted with an agent without emotion simulation. The findings indicated that participants perceived the emotionally expressive agent as more confident in communication. Participants also reported greater empathy towards the agent who simulated emotions than the agent who did not.

Existing studies have mainly explored emotional processing and recognition, particularly focusing on audio-video congruence as a crucial factor. The study of Torre *et al.* [2019] examined the manipulation of a multimodal agent's emotional expression, including "smiling face" and "smiling voice", as well as the agent's type (photorealistic or cartoon-like virtual human) to assess trust levels. The study employed a mixed-methods approach, combining behavioral data from a survival task, questionnaire ratings, and qualitative comments. The results highlighted the importance of emotional expressivity in the agent's voice, although its influence on trust behaviors was limited. Similarly, while participants rated the cartoon-like agent higher on various traits compared to the photorealistic one, the agent's style did not emerge as the most influential factor on trust behaviors. This study underscores the significance of employing a mixed-methods approach in human-machine interaction research, recognizing the contributions of both explicit and implicit perception and behavior to the success of the interaction.

In another study, Türkgeldi *et al.* [2022] seeks to explore the impact of familiarity with one's negotiation partner, specifically assessing the influence of opponent familiarity during human-agent negotiations. The study introduces a comprehensive human-agent negotiation framework and conducts a user experiment where participants engage in negotiations with avatars. These avatars either replicate the appearance and voice of a chosen celebrity or present unfamiliar characteristics. The results of the within-subject design experiment reveal that human participants tend to display greater collaboration when negotiating with a celebrity avatar they have positive feelings for, as opposed to a non-celebrity avatar.

Regarding the selection of the studies mentioned in this literature review, we employed a systematic process to identify and include research that is closely related to the central theme of our work. The papers were chosen based on their relevance to the topics of simulating emotions in virtual agents, the influence of emotions on trust, and the im-

part of these factors within the context of interactive games. Our selection process was guided by the goal of gathering insights and contextual information that would inform and complement our research. The chosen articles were considered valuable for providing a comprehensive background on the subject matter, offering diverse perspectives and findings from previous studies, and enhancing the theoretical foundation of our work.

Finally, our study stands apart from previous works by comprehensively presenting the development and validation process of a virtual agent's emotional expressions, specifically employed within a memory game set. Moreover, our research uniquely measures the effects of erroneous suggestions on human-agent dynamics, encompassing perception, trust, and task immersion. This approach sets our work apart by offering a detailed insight into the creation, validation, and practical implications of virtual agent emotions within the context of user interaction and trust.

3 Methodology

In this study, we employ a structured approach consisting of two interconnected phases. Each phase is thoroughly explored in dedicated subsections, offering a comprehensive understanding of the research objectives and encompassing distinct yet interdependent components. The initial phase, outlined in Section 3.1, focuses on the development and validation of a virtual agent known as Roboldo. This phase is required for establishing the groundwork, ensuring that participants accurately perceive the agent's emotions and expressions. Building upon the insights gained in the first phase, the second phase, detailed in Section 3.2, delves into the practical application of the virtual agent within a digital casual memory game. In this phase, we investigate how the agent's emotions and persuasive techniques impact participants' trust levels, perceptions, and overall gaming experience. This two-phase approach allows us to explore the nuanced interplay between the design of the virtual agent and its real-world influence, shedding light on the complex dynamics of human-agent interactions in gaming scenarios.

3.1 First Phase: Virtual Agent Development

3.1.1 Agent's Characteristics

A decision in achieving the project's goal is defining the characteristics of the virtual agent. Research has demonstrated that an agent's appearance can influence human interaction with it (Türkgeldi *et al.* [2022]; Torre *et al.* [2019]). Therefore, this project focused on defining the agent's appearance, shape, behavior, and unique traits. We gave the agent robotic features to ensure a neutral and inclusive representation, avoiding any specific gender representation (Shiban *et al.* [2015]). Additionally, we chose the name "Roboldo" with the specific intention of enhancing approachability and personability rather than implying a particular gender identity.

Initially, the agent used in this work was modelled. Modelling a virtual character can be done on paper or using computer software for drawing. However, in the first method of

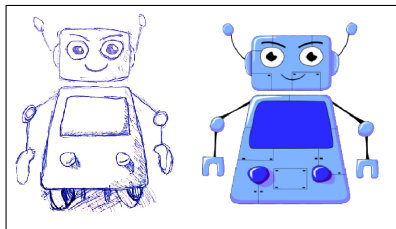
modelling, the agent needs to be digitized for use in the digital environment. In the second method, as the drawing is done using software, digitisation is unnecessary.

The agent's characteristics, such as appearance, shape, and peculiarities, are defined in this initial phase. Thus, it was decided that the agent would have the characteristics of a robotic being. This choice was made because the agent should not represent a specific gender (female or male), as studies have shown that virtual agents with such characteristics can influence the performance and motivation of participants (Shiban *et al.* [2015]).

Regarding the agent's modelling, a 2D drawing tool was used in this work. Thus, the 2D robotic agent was created using the "Inkscape" tool. This free software for 2D drawing provides numerous tools and supports vector images. Inkscape allows export to PNG images but also allows saving images in SVG, SVGZ, PDF, Postscript/EPS/EPSi, LaTeX (*.tex), POVray (*.pov), HPGL, and other formats (Inkscape [2020]). For this work, the "SVG" vector image format was chosen so that the character could be scaled and exported without loss of definition or image deformation.

In this way, a hand-drawn sketch of the agent was created, allowing the definition of its main characteristics. Once the agent's sketch was completed, it was drawn in Inkscape, where it was digitized, and the final artwork was applied. This involved adding colours and details to give the character more "life" and "charisma". Figure 1, on the left, shows the initial hand-drawn sketch, and on the right, the final artwork of the agent.

Figure 1. First sketch of the hand-drawn agent (left) and its final version (right).



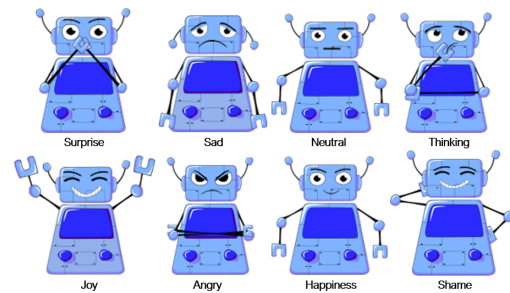
Source: Prepared by the authors.

3.1.2 Emotional Simulation

After determining the agent's appearance, we meticulously curated a set of emotional states for the agent to simulate, drawing upon the foundational research of Ekman [1971] and the insights from Disney's work on character expressions (Johnston and Thomas [1981]). The chosen emotional states included happiness, anger, sadness, neutrality, concentration, shame, joy, and surprise. Each emotion was carefully linked to specific arm positions, eye movements, mouth expressions, antenna adjustments, and body language. Our approach aimed to create simulations of easily perceivable emotions by combining both facial and body expressions, adhering to principles of character engineering and semiotics. Figure 2 visually represents the defined emotions.

To validate these simulations, a brief study was conducted in which participants had to identify the emotion the agent was expressing based on an image of the agent (de Lima

Figure 2. Roboldo simulating emotions.



Source: Prepared by the authors.

[2020]). This allowed us to assess the reliability of the drawn emotions and whether individuals could accurately identify the emotion being conveyed by the agent. To ensure the validity of the data collected in this study, rigorous statistical analysis was performed, accounting for potential biases or misinterpretations by participants. This analysis helped us obtain a more comprehensive understanding of the participant's ability to correctly identify the agent's simulated emotions and any potential sources of variation in their responses.

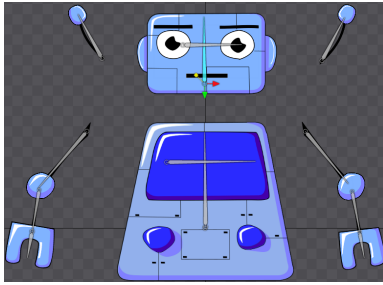
3.1.3 Animation Development

Following the validation of the emotions simulated by the agent, we implemented the animations that the agent performs to exert social influence and persuade the user. At this stage, we utilized the open-source 2D animation software, "DragonBonesPro"¹. We chose this tool because the developer had already worked with it and was aware of its operation and technological capabilities that meet the project's purpose. The DagonBonesPro employs the concept of animations with "Bones". This concept mimics the structure of the human body (Bones [2019]). Another advantageous feature of this tool is its seamless integration with the software used to develop the game scenario (further details in Section 3.2). The animations were designed to introduce dynamism to the character and prevent it from appearing static and lifeless, thus ensuring a higher level of immersion for the participant during the interaction. Furthermore, animations serve as a means for the agent to convey the emotions it "feels", thereby enhancing the player's perception of the emotion being simulated by the agent.

It is important to highlight that an animation was created for all expressions presented by Figure 2. The animation sequence remained consistent for all expressions simulated by the agent, commencing with the neutral expression and concluding with the expression representing the specific emotion. Thus, the animation transitions between these two expressions simulated by the agent. The character (agent) is divided into various parts to create the animation. These parts are imported and assembled within the tool to construct the agent's structure (as illustrated in Figure 3).

Then, the agent is divided into segments to create the animation; it is essential to determine a starting point, specifying the initial position for each part of the character in Frame 0 (zero). Additionally, it is necessary to define the final posi-

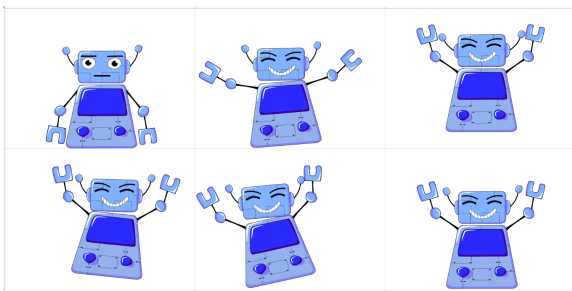
¹DragonBones Official Website. Available at: <https://dragonbones.github.io/en/index.html>. Accessed on: 8th November 2023.

Figure 3. Example of modelling in the DragonBonesPro tool.

Source: Prepared by the authors.

tion and frame. The tool already possesses the requisite information for generating the animation. The greater the distance between the initial frame and the final one, the slower the animation will appear.

This tool offers the flexibility to export animations in various formats, such as JSON and animated GIFs. It's worth noting that animations consist of sequences of images. For instance, the animation representing the emotion of joy comprises over 50 frames (images). In Figure 4, you can find a summarized depiction of the joy animation.

Figure 4. Summary of joy animation.

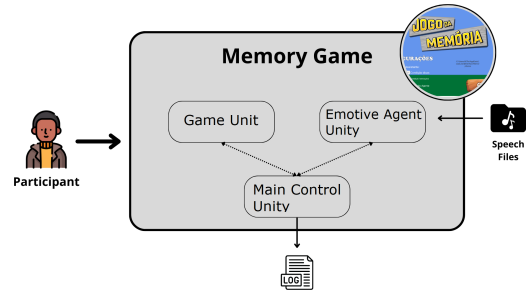
Source: Prepared by the authors.

As previously mentioned, this process was carried out for all of the agent's expressions, and the tool can export these animations as individual images (one image for each Frame) to the "Unity" Engine. Within Unity, these images are merged and transformed into unique animations for each emotion.

3.2 Second Phase: Game Scenario

For this study, a 2D casual game already familiar to most people, the "Memory Game", was developed. In this game, the participant's objective is to flip over two identical cards, and the game ends when the player successfully flips over all of the cards. A game with familiar mechanics was chosen to ensure participants did not waste time becoming familiar with the rules. The game was developed using a 2D and 3D game creation engine called Unity².

The memory game was divided into three units (Figure 5): (1) the "Main Control Unit", which manages communication and interaction between the units; (2) the "Game Unit", responsible for controlling the mechanics; and (3) the "Emotive Agent Unit", responsible for managing the agent's interactions and emotions (details in the following sections).

Figure 5. System architecture.

Source: Prepared by the authors.

The participant/player can interact with the game by choosing which cards to flip over in an attempt to make a pair. During this interaction, the game provides visual feedback (cards flipping over with each selection) and sound feedback (e.g., a "beep") to confirm the action. A file of "speeches" stores all the information for agent interaction, such as phrases and emotional states that the agent should express. A log file is responsible for keeping records of all interactions made in the game, whether made by the agent (speeches and emotions) or the player (selected cards).

3.2.1 Main Control Unit

This unit controls the interactions between the Game Unit and the Emotive Agent Unit. It detects the state of the game and the interactions made. Additionally, this unit is responsible for activating the Emotive Agent Unit and requesting emotions and phrases based on the player's actions. In summary, the Main Control Unit handles all communication. For example, when the participant chooses two cards, it verifies whether the selected cards match the ones indicated by the agent and reports to the Emotive Agent Unit whether the player followed the guidance. This unit also manages the log file, recording all interactions between the player and the agent.

3.2.2 Game Unit

In the game project, two screens are essential to achieve the goal of this research: the settings screen, which is only accessible by the game developer, and the main screen, where the cards and the agent are presented.

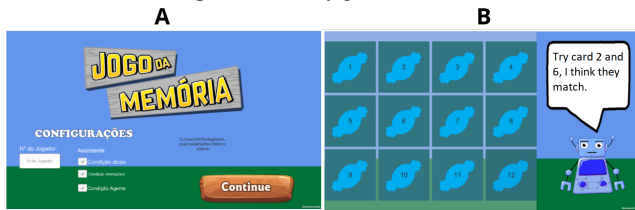
On the settings screen (Figure 6A), the necessary information for the game's correct functioning is provided and the scenario is configured according to the desired conditions. This information includes an identification created by the researcher for each participant/player, which is necessary for the log to identify each player's actions and the agent's condition, as it is essential to measure the agent's different behaviors.

On the game screen, the participant interacts with the developed system/game (Figure 6B³). The cards are marked with numbers to facilitate player identification and allow the agent to provide straightforward guidance/suggestions. During the game, all interactions made by the participant are recorded and saved in the log file.

²Unity Technologies. Available at: <https://store.unity.com/plans-individual>. Accessed on: 26th September 2023.

³The game was developed in Portuguese; however, in this paper, some images have been edited to have the text in English.

Figure 6. Memory game scenario.



Source: Prepared by the authors.

3.2.3 Emotive Agent Unit

During the game, the agent, positioned on the right side of the cards, assists the player by providing suggestions through animations, text, and speech. The game is designed to partially block when the agent interacts with the player, aiming to prevent the player from ignoring the agent and to encourage them to pay attention to the agent's actions or suggestions.

The Emotive Agent Unit is divided into three modules: the Agent Control Module, the Emotions Module, and the Voice Module.

The Agent Control Module controls the agent's response to player interactions, providing true or false instructions. For example, the agent may suggest that the player flip two cards that may or may not match, such as "Try flipping cards 2 and 3". If the player follows the agent's suggestion, the agent reacts positively with an animation expressing joy. Otherwise, an animation of sadness informs the player that the agent is disappointed with their decision. Figure 7 displays the game screen with the agent expressing happiness and a supportive message when the player accepts its suggestion.

The Emotions Module receives information from the Agent Control Module on whether or not the player followed the agent's advice and executes the corresponding animation based on the received data. This module also controls other animations, such as when the agent speaks and passive animations executed while the player is playing or thinking. These latter animations were programmed to give players a sense of immersion and make the agent appear lively rather than a static figure in the game environment.

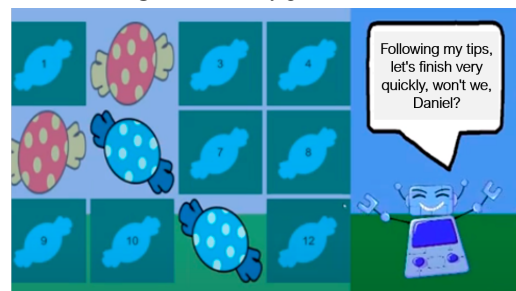
The Voice Module is responsible for managing the agent's voice. A Text-to-Speech software is used to reproduce the phrases received from the Agent Control Module as sound. This module also controls a dialogue balloon that presents the text of the spoken phrase from the Text-to-Speech. This process is necessary because if the player does not understand what the agent is saying, they can read the text of the speech in the balloon.

4 Research Methods

4.1 Ethical Considerations

We recognize the importance of adhering to ethical and moral norms in conducting research involving human beings. However, we faced an institutional limitation, as our university does not have a formally constituted Research Ethics Committee (CEP), and unfortunately, we were unable to submit our project to other Platform.

Figure 7. Memory game scenario.



Source: Prepared by the authors.

Despite this limitation, we want to emphasize our commitment to research ethics. During the execution of this study, we obtained informed consent from all participants and prioritized data privacy, confidentiality, and voluntary participation. Participants were informed that they could withdraw from the study at any time without facing negative consequences.

We understand that the absence of CEP involvement can be seen as a significant flaw. However, we would like to emphasize that our research was conducted with the utmost care and consideration for ethical principles. We are committed to learning from this experience and taking all necessary measures to ensure ethical compliance in future research.

4.2 Manipulation

In both phases, before starting the investigation, it was determined that there would be no age restrictions or requirements regarding familiarity with games. While our hypotheses do not directly involve the variable of age, we chose to collect this data because we believe it could be relevant for exploratory analyses or future research questions. Therefore, we decided to collect this data proactively.

For the second phase, we aimed to maintain a similar distribution of male and female participants in each study condition. Participants were divided into five (5) conditions, with an equal number of participants in each condition. The conditions are as follows:

- **C1:** Game without the agent's presence.
- **C2:** Game with the agent simulating emotions and providing correct suggestions.
- **C3:** Game with the agent not simulating emotions and suggesting correctly.
- **C4:** Game with the agent simulating emotions and providing false suggestions.
- **C5:** Game with the agent not simulating emotions and providing wrong suggestions.

These study conditions were created to understand whether the agent's presence, affective behavior (expression of emotion), and the accuracy of suggestions could influence trust, human-agent perception, following agent suggestions and the participant's experience in the task.

4.3 Procedures and Survey Design

The time interval between running the first phase and the second was one year, and participants could have participated in

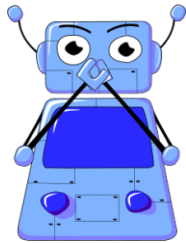
both studies. However, we did not carry out this verification, considering that we believe that participation in one phase would not affect the factors evaluated in the second phase.

4.3.1 First phase:

The questionnaires were available on the Google Forms platform. Each participant answered socio-demographic questions regarding age and gender. Furthermore, the questionnaire was created so that the participant saw an image of the agent simulating an emotion and then chose from multiple choices which feeling they thought the agent was expressing. In total, eight emotional expressions from the agent were presented to each participant. Figure 8 illustrates an example of a requested question.

Figure 8. Example of a question submitted to participants.

What emotion is being represented by the image below?



- Happiness
- Anger
- Sadness
- Neutrality
- Thinking
- Shame
- Joy
- Surprise
- Others..

Source: Prepared by the authors.

4.3.2 Second phase:

Conducted within university premises, this study primarily involved undergraduate students who voluntarily participated in the second phase. Participants were invited to participate voluntarily in the second phase of the study, with the indication that they would play a memory game and interact with a virtual agent (conditions with the presence of the agent). It was explained that the study would consist of three parts: a pre-questionnaire, an interaction with the game, and a post-questionnaire. After accepting the invitation, participants were shown the consent form, which outlined the purpose of the study, what their participation would involve, any potential risks or benefits, and their rights as participants, including the right to withdraw at any time without penalty. The study was explained in detail, and informed consent was obtained from each participant. Throughout the study, we adhered to ethical guidelines to ensure the protection of our participants' rights and welfare.

In the pre-questionnaire phase, we presented an image of the virtual agent, Roboldo, to the participants. This image served as the basis for participants to gauge their initial trust and perception of the agent. The specific questions asked in this phase will be detailed in the following section.

Following the pre-questionnaire, the researcher left the room, allowing the participants to interact with the agent through three rounds of a memory game. The choice of three rounds was made to avoid potential negative effects such as fatigue or boredom. After completing the three rounds, the researcher returned and provided the post-questionnaire. This questionnaire aimed to measure any changes in the participants' level of trust and perception towards the agent, as well as their level of immersion in the game.

Pre-Questionnaire and Post-Questionnaire

The questionnaires played a significant role in obtaining and analyzing data on the interactions. The questionnaires for this phase were developed and made available through the Google Forms platform. Participants who interacted with the agent were assigned to one of the following conditions: C2, C3, C4, or C5. Participants who did not interact with the agent were placed in condition C1.

The questionnaires aimed to collect data for the following aspects: (1) Demographic, (2) participants' experience with games (Poels *et al.* [2007]), (3) participant-agent perception (Rosenthal-von der Pütten *et al.* [2018]), and (4) the level of participant-agent trust (Benbasat and Wang [2005]).

We collected demographic data on participants' age and gender. We also gathered information on participants' gaming experience using a questionnaire inspired by the work of Poels *et al.* [2007]. According to the authors, the Game Experience Questionnaire has a modular structure consisting of the core questionnaire, the Social Presence Module, and the Post-game module. Due to its extensive nature, we focused on measuring the terms Negative and Positive Affect, which probe the fun and enjoyment of gaming. Participants responded using a 5-point unipolar intensity-based answering scale, ranging from "not at all" (0) to "extremely" (4). Additional questions were added to the pre-questionnaire, such as the number of hours played per week, preferred game style, and most played platforms. The questionnaire can be found in Supplementary Material link folder 1.

For the participant-agent perception (3), the Godspeed questionnaire (Bartneck *et al.* [2009]) was used. This questionnaire has five dimensions measuring specific human-agent perceptions: anthropomorphism, animacy, likability, perceived intelligence, and perceived safety. Participants indicated their responses using a 5-point Likert scale with bipolar options. The questionnaire can be found in Supplementary Material link folder 2.

According to Bartneck *et al.* [2009], anthropomorphism refers to attributing human form, characteristics, and behaviours to nonhuman agents. Animacy refers to assigning real properties to an entity by a user. Likeability is defined as the development of an agent's positive impression. Perceived intelligence is defined as the ability of an agent to adapt its behaviour to varying situations. Finally, perceived safety is the user's perception of the level of danger when in-

teracting with an agent and the user's level of comfort during the interaction. We did not measure the last dimension (perceived safety) since, in our study, the task performed by the agent did not represent a dangerous situation. The volunteer should point out, using a 5-point Likert scale with bipolar options, their perception regarding questions to measure the Godspeed dimensions.

To measure the level of participant-agent trust (4), the questionnaire presented in the work of Benbasat and Wang [2005] was used. This questionnaire evaluates trust in an agent across six dimensions: competence, benevolence, integrity, perceived usability, perceived ease of use, and intention to adopt. Participants responded on a 9-point Likert unipolar scale (strongly distrust - strongly believe), indicating their perception of the agent's activity. For this questionnaire, the dimensions that we believe are most related to the activity performed by Roboldo and the questionnaire creators recommend that for trust measurement are the first three dimensions. The questionnaire used can be found in Supplementary Material link folder 3.

4.4 Statistical Analysis

We used IBM SPSS V.28.0 packages and Microsoft Excel for the second study for the statistical analysis. The Shapiro-Wilk test Shapiro and Wilk [1965] was performed in our data set to identify the normality of data. As our data deviated from a normal distribution ($p > 0.05$), we applied the Wilcoxon signed-rank test to understand whether there was a difference in the level of trust and Godspeed perception between the conditions. In those tests, we considered 5% of statistical significance, i.e., p values of ≤ 0.05 .

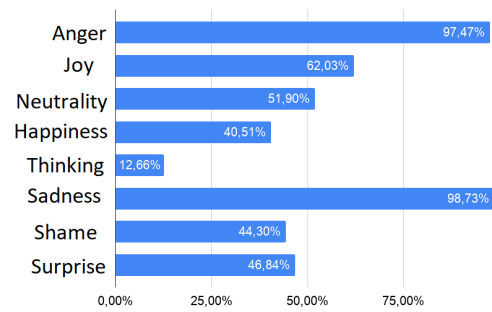
5 Results

5.1 First phase:

The questionnaire was available for 30 days, and a total of 79 people participated in the study by answering the questionnaire, with 59.5% (47) from the Female gender and 49.5% (32) from the Male gender, with an average age of 41, with a standard deviation of 13.64. The collected data is available at the supplementary material folder Virtual agent validation.

H1: After collecting the questionnaire responses, a basic statistical analysis was carried out, with only the percentages of the results obtained. The analysis considered how many participants agreed with the researchers' perception regarding the feeling that the agent is simulating in the image. For example, in Figure 8, the researchers believe that the agent is simulating the feeling of surprise, and 46.84% of the participants agreed with this representation. Figure 9 presents a graph with the proportion of agreement participants had with each emotion simulated by the agent. Through this graph, it can be seen that many expressions had a high degree of agreement, such as the expressions of Anger (97.47%), Sad (98.73%) and Joy (62.03%). Other expressions, such as Thinking (12.66%), Happiness (40.51%) and Shame (44.30%), had a lower agreement rate among participants.

Figure 9. Percentage of agreement by emotion.



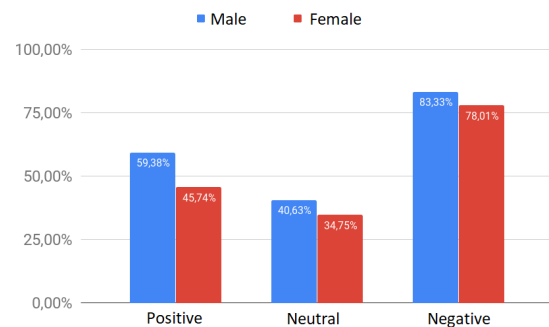
Source: Prepared by the authors.

H2: Our second hypothesis proposed a relationship between gender and agreement with the agent's expressions. The results of our analysis confirmed this hypothesis. We found that male participants agreed more with the agent's expressions.

In addition to the results obtained individually, we categorized emotions into Neutral, Negative, and Positive categories. Positive emotions included Joy and Happiness, Negative emotions included Anger, Sadness, and Shame, and Neutral emotions included Thinking, Surprise, and Neutral.

We found that there was greater agreement with Negative emotions, with Neutral being the category with the lowest percentage of agreement. These findings are illustrated in Figure 10, which shows the rate of agreement and disagreement for each of these categories, separated by the participant's gender.

Figure 10. Agreement based on emotion classification and participants' gender.



Source: Prepared by the authors.

5.2 Second phase:

This investigation was conducted in an isolated room with a random selection of people. A sample of 25 subjects (avg. 25 years old; $SD = 4.33$) participated and was assigned five per study condition (6 females (avg. 24.33; $SD = 5.12$) and 19 males (avg. 24.89; $SD = 4.03$)). The collected data is available at the supplementary material folder Memory game data.

H3: To investigate whether incorrect suggestions from the agent affect a person's level of trust, we first checked for any statistical differences in the data from participants who interacted with the virtual agent. We applied the Wilcoxon test to

the pre- and post-questionnaire data and found significant differences in the perceived competence ($Z = 2.092, p = .036$) and benevolence ($Z = 2.175, p = .030$) factors.

We then divided the data according to the conditions created for this phase of the study and applied the Wilcoxon test again. Interestingly, only condition C4 (where the agent simulated emotions and provided false suggestions) showed statistical differences in the competence ($Z = 2.032, p = .042$) and benevolence ($Z = 2.023, p = .043$) factors before and after the interaction. No statistical differences were found in the other conditions.

H4: To investigate whether false information provided by the agent can affect a person's perception of this agent (H4), we analyzed the participants' responses from the Godspeed questionnaires in the conditions with the agent's presence (C2-C5). We applied the Wilcoxon test to the pre- and post-questionnaire data and found no significant differences in the evaluated dimensions of anthropomorphism ($Z = -0.825, p = .409$), animacy ($Z = -1.585, p = .113$), likeability ($Z = -1.121, p = .262$), and perceived intelligence ($Z = -0.261, p = .794$).

We then divided the data according to the study conditions. Our results suggest that participants had a different perception of the dimensions of perceived intelligence in C2 ($Z = -2.041, p = .041$) and C3 ($Z = -2.023, p = .043$) and anthropomorphism ($Z = -2.000, p = .046$) in C4. In C5, we did not find any statistical differences.

H5: To investigate whether people with a level of trust that was not negatively affected by the agent's false suggestions follow the tips provided by the agent more (H5), we analyzed the game log file. This file recorded the interaction information of the player, Agent, and game. We observed the number of interactions and the correlation between the rate of agreement and disagreement. In other words, we analyzed the number of times participants followed or did not follow Roboldo's tips.

Our analysis revealed that the number of interactions with the agent was much higher in conditions C4 (302) and C5 (290) compared to conditions C2 (93) and C3 (97). Interestingly, the number of interactions without the agent (C1 192) was greater than the number of interactions with the agent telling the truth (C2 and C3).

We also found that the number of interactions where participants agreed with Roboldo was greater in conditions C4 (109) and C5 (87) than in conditions C2 (75) and C3 (70). The difference in the number of interactions where participants did not agree with the agent was much greater between conditions C4 (193), C5 (203) and C2 (18), C3 (27). These findings are summarized in Table 1. The first column shows the five conditions of the game, as explained in section 4.2. The second column shows the total number of interactions that occurred in each condition, where an interaction is defined as a pair of cards selected by the participant. The third and fourth columns show the number of times that the participant agreed or disagreed with the agent's suggestion, respectively. Agreement means that the participant followed the agent's suggestion and flipped the cards indicated by the agent. Disagreement means that the participant ignored the agent's suggestion and flipped other cards. These numbers indicate the level of trust and compliance that the participants

had towards the agent in each condition. NA stands for Not Applicable, as condition C1 did not have the agent's presence.

Table 1. Number of interactions, Agreement and Disagreement.

	N Interaction	Agreement	Disagreement
C1	192	NA	NA
C2	93	75	18
C3	97	70	27
C4	302	109	193
C5	290	87	203

The analysis regarding the number of interactions between the player and the agent was also carried out according to the rate of agreement and disagreement between the three Rounds (3 matches) of the game. Table 2 shows the number of interactions, agreements, and disagreements for each condition and round. The results indicate that in the conditions where the agent helped the player (C2 and C3), the number of interactions and agreements remained stable across the rounds. However, in the conditions where the agent hindered the player (C4 and C5), there was a decrease in the number of interactions and agreements in each round.

Table 2. Number of interactions, Agreement and Disagreement per round.

C	Round	N Interaction	Agreement	Disagreement
C2	R1	31	25	6
	R2	32	25	7
	R3	30	25	5
C3	R1	32	25	7
	R2	31	25	6
	R3	34	20	14
C4	R1	133	63	70
	R2	84	21	63
	R3	85	25	60
C5	R1	123	66	57
	R2	109	17	92
	R3	58	4	54

H6: To investigate whether the presence of an agent within the game positively affects the player's experience if it provides accurate advice (H6), we analyzed the responses from the game experience questionnaire. We applied the Wilcoxon test to the data and found statistical differences for the condition without Roboldo (C1) for negative ($Z = -2.023, p = 0.043$) and positive affects ($Z = -2.032, p = 0.042$).

We then divided the data according to the study conditions and applied the Wilcoxon test again. For condition C1 (without the virtual agent), there were differences according to the Wilcoxon test for negative affect ($p = .043, Z = -2.023$) and positive affect ($p = .042, Z = -2.032$). Regarding the conditions with the virtual agent, the following differences were obtained: C2, only positive affect had statistical differences ($p = .042, Z = -2.032$), whereas no differences were found for negative affect ($p = .225, Z = -1.214$); Similar behaviour was observed for C3, where only positive affect had statistical differences ($p = .043, Z = -2.023$), whereas no differences were found for negative affect ($p = .686, Z = -.405$); C4, only positive affect had statistical differences

($p = .046$, $Z = -1.997$), whereas no differences were found for negative affect ($p = .498$, $Z = -.677$); C5, only positive affect had statistical differences ($p = .042$, $Z = -2.032$), whereas no differences were found for negative affect ($p = .136$, $Z = -1.490$).

6 Discussion

6.1 First phase:

H1: The results obtained in this study suggest that people did not fully perceive the positions of the agent's body elements as being the emotions planned by the researcher. This shows the importance of validating the agent's emotions before using them in a study. Furthermore, it was observed that some simulated expressions raised doubts in the participants, such as the difference between happiness and joy.

H2: Our observations revealed a gender difference in the gender's perception of emotions simulated by the agent. Specifically, male participants were more likely to concur with the emotions instigated by the researcher than female participants. This could potentially be attributed to the fact that the animations and emotion simulations of the agent were developed by a male researcher.

Our decision to focus on the gender difference in emotion perception was driven by its potential implications on the design and evaluation of virtual agents for different user groups. These findings are in line with existing literature that suggests men and women may perceive facial expressions differently (Montagne *et al.* [2005]). Moreover, more recent studies such as Fischer *et al.* [2018]; Cameron *et al.* [2018]; Ghazali *et al.* [2018]; Nomura [2017] provide further support to our findings in robotic agent facial expressions.

However, we acknowledge that these results are preliminary. Further research involving a larger and more diverse sample is necessary to substantiate these findings.

6.2 Second phase:

H3: The level of trust in the agent was notably affected by the specific conditions to which participants were exposed. A noticeable decrease in Roboldo's perceived competence and benevolence was observed, particularly in condition C4, where the agent provided misleading clues, resulting in no card matches. This decrease aligns with the *a priori* level of trust formed solely from the agent's initial presentation. However, when false hints were given, this trust significantly reduced as the agent transitioned from being perceived as helpful to hindering the task. While C5 also provided incorrect information, the lack of emotional expression differentiated it from C4. This highlights that incorrect suggestions impact trust, especially when accompanied by emotional cues.

H4: Findings regarding the person-agent relationship perception revealed apparent shifts in participants' evaluations before and after interaction based on the agent's behavior. In conditions where accurate hints were provided, participants reported a significant change in their assessment of the agent's intelligence. The agent's ability to aid players in quickly completing the game influenced the perception

of its intelligence positively. In the condition where false hints were given along with emotional simulation (C4), there was a distinct shift in the perception of the agent's anthropomorphism post-interaction, with participants considering the agent as more 'alive'. The lack of differences in some dimensions may be attributed to the limited sample size in each condition, underscoring the need for a more extensive sample to comprehensively observe human-agent perception across the dimensions of the Godspeed questionnaire.

H5: The analysis of Log Files highlighted variations in participants' responsiveness to the agent's guidance, notably influenced by the veracity of the provided tips. Observations revealed a distinct pattern in the number of interactions within different conditions. Notably, fewer interactions were recorded when the agent provided accurate tips, as players adhered closely to Roboldo's guidance, resulting in fewer game interactions and smoother gameplay. Conversely, in conditions with false tips, the interactions surged, as participants initially followed the agent's suggestions, assuming their accuracy. Moreover, participants exhibited an interesting trend regarding the number of interactions per round in C4 and C5. There was a decrease in interactions with each subsequent round, indicating a quick realization among participants that the agent's guidance was erroneous, leading to a subsequent disengagement from Roboldo's misleading suggestions.

H6: Statistical analysis revealed significant differences in the game experience, specifically in positive and negative affects across all conditions. When examining participants' interactions with the agent before and after the game, noticeable changes were detected in positive and negative affects. Positive affects were notably different in the second and third study conditions, whereas in C4 and C5, negative affects showed distinctions. These findings suggest that the agent's provision of accurate or false clues significantly impacts players' game experience. However, further studies with a larger sample size are imperative to consolidate and fortify these findings.

6.3 Threats to Validity and Limitations

Although our study provides valuable insights into human-agent interaction, we need to acknowledge potential factors that might affect the validity of our results. The sample size we used was relatively small. This poses a threat to the external validity of our findings, which limits their generalizability. In addition, the emotional expression design of the agent may not fully capture the nuances of emotions. This could potentially introduce ambiguous interpretations for the participants and pose a threat to the internal validity of our study. The structured experimental conditions we used were designed to simulate trust scenarios. However, they may only partially mirror real-world situations. This could affect the conclusion validity of our study because our findings may not fully align with the dynamics observed in more naturalistic settings. Moreover, both the limitations of the sample size and the emotional expression design contribute to potential threats to construct validity. To address these threats, we need future research endeavors with larger and more diverse samples, refined emotional expression designs, and studies conducted in naturalistic settings. These efforts will enhance

the overall validity and robustness of our findings.

7 Conclusion

This research delved into the dynamics of virtual agents their impact on human interaction, and decision-making. Firstly, the study underscored the need to validate and refine agent emotions before deployment. Moreover, the research illuminated valuable insights, particularly in the role of agent performance on user trust and experience. It highlighted the influence of deliberate errors in the agent's responses on users' confidence, perception, and overall task experience.

The study's first phase presented insights into the perception of emotions simulated by virtual agents. While the findings did not entirely align with the anticipated perception of specific emotions through the agent's body positions, it emphasized the need to validate and refine agent emotions before their use in studies. Furthermore, the study revealed a gender difference in emotion perception, with male participants more likely to align with the intended emotions than female participants. These gender-based distinctions in perceiving facial expressions correlate with existing literature and recent studies in robotic agent facial expressions. Nevertheless, recognizing the preliminary nature of these results, further extensive and diverse research is imperative to solidify these initial findings. This contributes to a more comprehensive understanding of gender-specific emotion perception and offers an understanding for designing and evaluating virtual agents across diverse user groups.

The second study underscored the critical role of virtual agent performance in user trust and experience. Our findings demonstrated that deliberate errors in the agent's responses had a detrimental effect on users' confidence, perception, and overall task experience. Conversely, when the agent provided helpful tips and information, users' experiences were significantly improved.

However, the study's outcomes may have been influenced by the relatively small sample size and limitations in simulating the agent's emotions. To build on these insights, we recommend future studies with larger sample sizes and developing agents with enhanced expression capabilities to further improve user accuracy and connection. This involves exploring various methodologies or design adaptations to enhance the agent's emotional range within existing technological frameworks.

In essence, this research contributes to a deeper understanding of how virtual assistants can shape human decision-making, trust, and engagement. The findings can guide the development of more effective virtual assistants, ultimately enriching the user experience. We aspire that this work paves the way for the design of virtual agents that provide invaluable assistance and foster greater user trust and engagement.

Declarations

Acknowledgements

We would like to inform you that the present work is an extended and revised version of a previous research paper published by DOI:

<https://doi.org/10.5753/wics.2023.229689>

Authors' Contributions

RP: Conceptualization, Writing – Original Draft, Writing – Review and Editing, Project Administration. AA: Writing – Review and Editing. DT: Conceptualization, Methodology, Investigation. AL: Software, Visualization. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

https://osf.io/qr8vk/?view_only=b4099caae9944c24b4ce35c7613a0717

References

- Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81.
- Benbasat, I. and Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the association for information systems*, 6(3):4.
- Bones (2019). After all, what is rigging and how does it help with character creation? <https://saga.art.br/afinal-o-que-e-rigging-e-como-ajuda-na-criacao-de-personagens/>.
- Cameron, D., Millings, A., Fernando, S., Collins, E. C., Moore, R., Sharkey, A., Evers, V., and Prescott, T. (2018). The effects of robot facial emotional expressions and gender on child–robot interaction in a field study. *Connection science*, 30(4):343–361.
- Chowanda, A., Flintham, M., Blanchfield, P., and Valstar, M. (2016). Playing with social and emotional game companions. In *International Conference on Intelligent Virtual Agents*, pages 85–95, Springer. Cham.
- Cicirelli, G., Marani, R., Petitti, A., Milella, A., and D'Orazio, T. (2021). Ambient assisted living: a review of technologies, methodologies and future perspectives for healthy aging of population. *Sensors*, 21(10):3549.
- Correia, F., Guerra, C., Mascarenhas, S., Melo, F. S., and Paiva, A. (2018). Exploring the impact of fault justification in human-robot trust. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems*, pages 507–513.
- Cuadra, A., Li, S., Lee, H., Cho, J., and Ju, W. (2021). My bad! repairing intelligent voice assistant errors improves interaction. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–24.
- Darwin, C. and Prodger, P. (1998). *The expression of the emotions in man and animals*. Oxford University Press, USA.
- de Lima, A. T. (2020). *Influência de Agente Virtual Afetivo nas Decisões do Jogador em um Jogo Casual*. (Influence

- of *Affective Virtual Agent on Player's Decisions in a Casual Game*, page 26–35. State University of Rio Grande do Norte.
- Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation*. University of Nebraska Press.
- Fischer, A. H., Kret, M. E., and Broekens, J. (2018). Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis. *PloS one*, 13(1):e0190712.
- Gass, R. H. and Seiter, J. S. (2018). *Persuasion: Social influence and compliance gaining*. Routledge.
- Ghazali, A. S., Ham, J., Barakova, E. I., and Markopoulos, P. (2018). Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers in Robotics and AI*, 5:73.
- Guzman, A. L. and Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New media & society*, 22(1):70–86.
- Hancock, P., Billings, D., and Schaefer, K. (2011). Can you trust your robot? *Ergonomics in Design: The Quarterly of Human Factors Applications*, 19:24–29. DOI: 10.1177/1064804611415045.
- Hashemian, M., Paradedda, R., Guerra, C., and Paiva, A. (2019). Do you trust me? investigating the formation of trust in social robots. In *EPIA Conference on Artificial Intelligence*, pages 357–369, Springer. Cham.
- Inkscape (2020). Frequently asked questions — for inkscape users. <https://inkscape.org/pt-br/aprender/faq/>.
- Johnston, O. and Thomas, F. (1981). *The illusion of life: Disney animation*. Disney Editions, New York.
- Lewis, M., Sycara, K., and Walker, P. (2018). The role of trust in human-robot interaction. *Foundations of trusted autonomy*, pages 135–159.
- Montagne, B., Kessels, R. P., Frigerio, E., de Haan, E. H., and Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity?. *Cognitive processing*, 6(2):136–141.
- Moyer-Gusé, E. (2008). Toward a Theory of Entertainment Persuasion: Explaining the Persuasive Effects of Entertainment-Education Messages. *Communication Theory*, 18(3):407–425.
- Nomura, T. (2017). Robots and gender. *Gender and the Genome*, 1(1):18–25.
- Poels, K., de Kort, Y., and Ijsselstein, W. (2007). D3. 3: game experience questionnaire.
- Ragni, M., Rudenko, A., Kuhnert, B., and Arras, K. O. (2016). Errare humanum est: Erroneous robots in human-robot interaction. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 501–506.
- Rosenthal-von der Pütten, A. M., Krämer, N. C., and Herrmann, J. (2018). The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics*, pages 1–14.
- Salem, M., Eyssel, F., Rohlfing, K. J., Kopp, S., and Joublin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, 5:313–323.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Shiban, Y., Schelhorn, I., Jobst, V., Hörnlein, A., Puppe, F., Pauli, P., and Mühlberger, A. (2015). The appearance effect: Influences of virtual agent features on performance and motivation. *Computers in Human Behavior*, 49:5–11.
- Timotheou, S., Miliou, O., Dimitriadis, Y., Sobrino, S. V., Giannoutsou, N., Cachia, R., Mones, A. M., and Ioannou, A. (2023). Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. *Education and information technologies*, 28(6):6695–6726.
- 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*, pages 1–6.
- Türkgeldi, B., Özden, C. S., and Aydoğan, R. (2022). The effect of appearance of virtual agents in human-agent negotiation. *AI*, 3(3):683–701.
- Yang, Y., Ma, X., and Fung, P. (2017). Perceived emotional intelligence in virtual agents. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2255–2262.