Developing a Scale for Measuring the Believability of Virtual Agents

Siqi Guo[†]   Nicoletta Adamo[‡]   Christos Mousas[§]

Department of Computer Graphics Technology, Purdue University, West Lafayette, IN 47907, U.S.A.

Abstract

Creating believable virtual agents has long been the focus of artists and scientists. A believable agent enables the audience to be emotionally involved in a narrative, willingly suspending their disbelief for the pleasure of appreciating literacy, drama, film, etc. In the past few decades, believability has become the goal of virtual reality developers and researchers. In the realm of virtual reality, it is commonly accepted that a believable virtual agent should have personality, emotion, agency, intelligence, and more. Despite its seemingly complicated requirements, believability is a fragile product that can be easily jeopardized by missing one or more of these elements. In this paper, we review the questionnaires that past researchers have used on the topic of virtual agents' believability. Based on the prevailing questions identified in the relevant studies, we propose a scale—aiming at standardizing one—for measuring the believability of virtual agents. We recommend that future research involving virtual agents refer to this scale to evaluate the level of their believability.

CCS Concepts
• ***Human-centered computing** → **Virtual reality; User studies;***

1. Introduction

In traditional story arts, a believable character enables the audiences' "willing suspension of" disbelief [Col71]. In the field of animation, games, and virtual reality, how to accurately define the believability of a virtual character/agent is still an open question [TYKS12]. It is generally accepted that believability refers to the extent to which people believe that the agent they observe and interact with is an autonomous, sentient being [LS97]. Believability does not equal realism [Mat01]. Realism contributes to the believability of a virtual agent; however, believability is a multidimensional term for personality, emotion, agency, intelligence, etc. These dimensions work together to bring the audience an "illusion of life" [TJT95].

Believability can be divided into character believability and player believability. Character believability refers to the extent to which the viewer believes the character is a real being, while player believability is the degree to which the player believes a real human controls the character [TYKS12]. In this study, the term "believability" specifically denotes the virtual agents' character believability. In the meantime, a standardized way of measuring virtual agents' believability has yet to be proposed [CAC*22]. Thus, considering that believability dimensions differ in prior research, in the following section, we identify a list of criteria for believable virtual agents that were commonly adopted by previous studies and organize them roughly in the order of lower level or static to higher level or dynamic.

2. Believability Dimensions

2.1. Visual Properties

The appearance of virtual characters influences human perception [ZZM19]. Bogdanovych et al. [BTS16] argued that a believable appearance is significant to the overall believability of a virtual agent since it impacts the users' first impression of the agent. Hetherington [Het15] categorized appearance into facial and body appearance in measuring the believability of hyper-realistic characters, and the author also mentioned the problem of an uncanny valley associated with the appearance of virtual agents. Another appearance-related problem is that the visual design of a virtual agent helps set the user's expectations toward its behavior and capacity; therefore, perceptual believability can be jeopardized when there is a mismatch [GKK18]. Also, a study by Lam et al. [LCM*23] suggested that, when corresponding with voice pitch, the visual properties influence the virtual character's perceived believability. Other than aesthetic considerations, visual impact, a concept first proposed by Lester and Stone [LS97], should be included in the dimensions of believability, which refers to the amount of human attention held by an agent [GPMJ13].

[†] guo477@purdue.edu
[‡] nadamovi@purdue.edu
[§] cmousas@purdue.edu

2.2. Behavior

The behavior of the virtual agent, including its coherence and understandability, has since been widely adopted by studies as a dimension in measuring believability, following its proposal by Gomes et al. [GPMJ13]. The significance of behavior has its roots in the nature of virtual agent believability, as people are able to see a virtual agent's demonstrated behaviors but not the inner states that drive them. Researchers commonly agree that a believable virtual agent's behavior should be coherent [GPMJ13], natural, and appropriate to the context [DNP11, BCR17], predictable [GPMJ13] but yet has unpredictabilities [BRDL17] that make them more human-like.

2.3. Awareness

Awareness believability is an essential part of the overall believability of a virtual agent [BTS16]. A virtual agent needs awareness for its transition from an acting character to a reacting or interacting agent. As suggested by Ijaz et al. [IBS11], awareness is a broad concept that includes environmental awareness, self-awareness, and interaction awareness. Environmental awareness refers to the virtual agent's ability to gain and make use of information about the dynamic environment through perception. Interaction awareness is closely related to the virtual agent's ability to form social relationships, and it requires the virtual agent to be at least partially aware of the structures of human interactions [DOQ02]. Whereas environment and interaction awareness are mainly based on the virtual agent's perception of its surrounding world, self-awareness emerges from its agency [MHC21] as well as the focusing of attention on itself [Mor11].

2.4. Social Relationships

Our daily lives revolve around interactions with other humans. For virtual agents, their human-likeness also depends on their ability to form and maintain social relationships with humans or other agents in cases where there is more than one agent [AP09, AWFM23, GPS*20]. The prerequisite for social relationships is social awareness [BYA*12]. A virtual agent should first be aware of the existence of others to engage in social activities. In the realm of games, such social activities usually revolve around communicating and coordinating with players or other agents [Liv06]. Warpefelt et al. [WJV13] suggested the importance of the appropriateness of the virtual agent's social behaviors within the context.

2.5. Intelligence

Whether or not the perceived intelligence of a virtual agent meets the user's expectations can affect its believability. Psychologists generally accept that there are four types of intelligence: intelligence quotient, emotional quotient, social quotient, and adversity quotient. The first two are widely studied as predictors of success in different population groups. In measuring intelligence, we focus on the general intelligence (intelligence quotient) instead of emotional intelligence, which measures one's emotional, personal, and social aspects of intelligence [DH00] and should be a separate dimension of believability [Loy97, AP09, GPMJ13]. The intelligence quotient usually refers to one's conceptual and analytical intelligence [ALA*05]. For virtual agents, intelligence is conveyed through their ability to plan, learn, and adapt. Further, the ability to hold memories is one of the basic characteristics of an artificial intelligence (AI) system's intelligence quotient [LSL17].

2.6. Emotion

The believability of a virtual agent can be enhanced by demonstrating emotional behaviors, which include expressing socially appropriate, easy-to-understand emotions in reaction to a stimulus [DNP11, GPMJ13, TC, SRM23]. Developing computational models that give the virtual agent the capacity to express believable emotions has been the focus of recent research [RPCP16]. The emotions of virtual agents are mainly conveyed through their facial and bodily expressions and voice tones [HHC18]. Basic or prototypical emotions investigated by researchers on emotional virtual agents are joy, sadness, fear, surprise, and anger [HHC18]. Recent studies have also argued for the importance of empathy in virtual agents, which refers to the ability to understand and respond to others' emotional states [PLBW17].

2.7. Personality

Personality is one of the most fundamental requirements for a virtual agent to be believable [DNP11]. A virtual agent's personality is its pattern of behavior and its way of interacting with the surrounding environment [BYA*12]. As suggested by McRorie et al. [MSM*11], such patterns should remain coherent for an extended period when users interact with the same virtual agent in multiple scenarios to maintain their believability. To describe personality traits, the five-factor model is often used, which suggests that the five basic dimensions of personality traits are neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness [MCJ97]. Researchers of virtual agents have also adopted the model to measure their personalities [PP05, CVB*12].

2.8. Agency

Given the limited research that includes questions on the virtual agent's agency and self-awareness in measuring believability, this study argues for their necessity. Agency and self-awareness are two closely interconnected terms. For ease of understanding, this study treats them as two parallel concepts in identifying the dimensions of believability. Agency is a psychological term that refers to the sense of generating one's behaviors, and it is the key to self-awareness [MHC21], while self-awareness is a much more abstract term that represents the capacity to focus attention on oneself [Mor11]. Ijaz et al. [IBS11] suggested in a more practical way of quantifying a virtual agent's self-awareness that possessing goals, plans, beliefs, and being aware of the current scene the virtual agent is situated in and its position within it are elements that contribute to self-awareness. Avradinis et al. [APA13] argued that, to a great extent, believability comes from the virtual agent's perceived autonomy, and it is the prerequisite of believability. Riedl and Stern [RS06] suggested that autonomy in a virtual agent can enhance its believability.

One major element of agency is the possessing of goals, or being

self-motivated [PHN*09]. A believable virtual agent should give users the illusion that they are observing and interacting with an autonomous being, acting based on its desires [LS97, DS03]. As also suggested in Loyall's [Loy97] requirements for believability, the "appearance of goals" and "concurrent pursuit of goals and parallel action" contribute to a virtual agent's "illusion of life" [TJT95].

2.9. Believability as a Multidimensional Concept

Several researchers that used questionnaires had questions on the virtual agent's overall believability, such as "The reaction of the agent was believable." [vdHBR14] and "The character was a believable real estate agent" [RB14]. We argue that such general questions should be included in a standardized questionnaire since believability does not come from a single aspect; instead, it arises from the combination of multiple interconnected dimensions [PHN*09], while it can also be easily jeopardized by the absence of one or more features.

3. Toward a Standardized Believability Scale

Similar to the definition and requirements, the assessment of believability is also open to debate [IBS11, CAC*22]. Previous studies have adopted both qualitative and quantitative methods to measure the believability of virtual agents. Commonly used techniques include subjective methods, such as questionnaires and interviews, and objective methods, such as recording participants' physiological responses [TYKS12].

Despite the disadvantages associated with its subjectiveness, questionnaires are the most widely used measurement of believability. Popular forms of questionnaire questions are Boolean, ranking, and preference [TYKS12]. Compared to Boolean and preference questions, the ranking questions do not limit the participants' responses to binary answers (e.g., yes or no); instead, they generate responses that not only show the tendency (e.g., I think the agent is/is not believable) but also indicate the level of confidence toward their choices through rating. Therefore, it is reasonable to use scales with ranking questions in experiments to reveal the believability of virtual agents.

In contrast to fast-growing research on the topic of believability, there is not yet a standardized questionnaire measuring the believability of virtual agents. To the best of our knowledge, many of the available believability questionnaires were adapted from the template proposed by Gomes et al. [GPMJ13]. Some studies have also combined questions from the Godspeed questionnaires [BKCZ09], the Believability Index [GTBH06], and the Dimensions of Mind Perception [GGW07], while others have developed their own questionnaires. The various versions of the questionnaires pose significant challenges to comparing the experimental results of different studies, which urged the development of a standardized questionnaire in measuring believability.

In this paper, we organize and classify the questions used in previous experiments that measure believability in various scenarios. We found that some dimensions of believability (e.g., personality) were associated with a relatively limited number of questions from the believability questionnaires. To address their significance, we

also reviewed experiments from the respective fields of study to obtain frequently used questions. Finally, we propose a scale to be used in future studies. Designing a believability scale poses multiple challenges caused by the different designs of experiments and different hardware setups across different research labs. A standardized believability scale will help researchers more efficiently compare the level of believability across different setups. It will also help us to understand better which factors play important roles in believability and how to achieve believability when developing intelligent virtual agents.

3.1. Review

In the first stage of this study, we collected 25 scholarly articles that adopted or proposed questionnaires relevant to believability through a systematic search in the disciplines of human-computer interaction (HCI), human-robot interaction (HRI), and animation. Specifically, we extensively searched databases, including IEEE Xplore Digital Library, ACM Digital Library, SpringerLink, and ScienceDirect. Also, we focused on the studies after 1997, when Loyall [Loy97] proposed a comprehensive list of requirements for developing believable agents and thus laying the foundation for the assessment of believability. Table 1 provides a summary of the studies reviewed. It also details the specific types of agents each study addressed and the scale used when a questionnaire was utilized. This categorization aids in understanding the primary focus and context of each research at a glance.

Then, we adopted a set of inclusion and exclusion criteria and filtered out twelve studies for the next stage of the reviewing process. The inclusion criteria were: complete studies on the human perception of virtual agents, non-player characters, or robots; studies yield experimental results on the perceived believability and adopted questionnaires as a method in accessing the believability of virtual agents, non-player characters, or robots; or studies that proposed methods of measuring believability without conducting experiments. Specifically, this includes a template of the believability questionnaire [GPMJ13], the believability index [GTBH06], and the criteria for believable game AI [Liv06]; and studies published on proceeding-based conference paper or peer-reviewed journals. The exclusion criterion was the questionnaires exclusively utilized scalar evaluations, prompting participants to indicate their perceptions on a scale from 1 to 7. While this method definitely yields valuable data on the subject matter explored, we excluded them for limited granularity concerning the topic we are investigating.

In the subsequent stage, we classified all the questions sourced from the questionnaires adopted by the selected studies into nine main categories:

- **Visual properties (or appearance) of the virtual agent.** This has been used to assess the level of realism of an agent or determine whether the visual design of the agent is coherent with its surrounding environment.
- **Behavior.** This concept is typically employed to measure both the believability of the motion when the virtual agents move and their behavior pattern in general.
- **Awareness.** Virtual agents' awareness of the surrounding environment and their interactions with humans or other agents can be measured with the awareness concept.

Table 1: *The reviewed papers at a glance, listed by year, then in alphabetical order.*

Citation	Year	Types of Agents					Measurement Scale
		VA	Game NPC	Animated Character	Robot	Generic Agent	
Gorman et al. [GTBH06]	2006		•				5-point Likert
Bosse and Zwanenburg [BZ09]	2009	•					7-point Likert
Poel et al. [PHN*09]	2009				•		5-point Likert
Demeure et al. [DNP11]	2011	•					7-point Likert
Ijaz et al. [IBS11]	2011	•					5-point Likert
Gomes et al. [GPMJ13]	2013	•					N/A
Ham et al. [vdHBR14]	2014	•					5-point Likert
Mitarai and Umemuro [MU14]	2014					•	7-point Likert
Richards and Bransky [RB14]	2014	•					4-point Likert
Bogdanovych et al. [BTS16]	2016	•					5-point Likert
Barreto et al. [BCR17]	2017		•				5-point Likert
Bevacqua et al. [BRDL17]	2017	•					6-point Likert
Hetherington [Het15]	2017			•			9-point Likert
Rodrigues and Martinho [RM17]	2017	•					N/A
Ishihara [III*18]	2018		•				5-point Likert
Pacheco et al. [PTPL18]	2018		•				10-point Likert
Coninck et al. [DCYSV19]	2019	•					N/A
Tsiourti et al. [TWWV19]	2019				•		5-point Likert
Guimaraes et al. [GPS*20]	2020	•					5-point Likert
Daylamani-Zad et al. [DZA20]	2021		•				5-point Likert
Dickinson et al. [DJC*21]	2021	•					N/A
Even et al. [EBB21]	2021		•				Binary scale
Tjokrosetio et al. [TC]	2021		•				4-point Likert
Curtis et al. [CAC*22]	2022			•			7-point Likert
Silva et al. [SRM23]	2023		•				9-point Likert

- **Social relationships (with humans or other agents).** This element is assessed whenever the human interacts with the virtual agent during the experiment. In scenarios with more than one virtual agent, questions on the social relationships between the virtual agents are also often asked.
- **Intelligence.** As mentioned in previous sections, one of the key elements of virtual agents' believability is the "illusion of life." Intelligence and agency are the two main factors that contribute to this concept. Questions on intelligence are often presented in relatively complicated scenarios in experiments in which the agent demonstrates the capacity for planning and memorizing.
- **Emotion.** This captures how the virtual agent perceives human emotions, as well as when the agent expresses emotions.
- **Personality.** This is often used in questionnaires when determining whether or not the virtual agent has a personality and what their types of personality are.
- **Agency.** This is employed when assessing the virtual agent's autonomy. These questions are usually asked in the form of whether the agent has personal goals.
- **Overall believability.** A popular element in most experiments, overall believability questions are usually asked in the form of rating believability or human likeness.

This method of classification is generally consistent with a previous study on the believability of virtual agents [EK17], which summarized the dimensions for believable AI, game non-player characters, and virtual characters proposed in early studies, provided a comprehensive guide to the dimensions of believability for different forms of agents prevails in interactive media [Mat01, Loy97, GPMJ13, LH15, BTS16]. Specifically, in Table 1 in the supplementary material document, we outline the elements of believability for virtual agents, robots, and creatures from research since 1997, based on ElSayed and King's literature review [EK17]. The literature indicates that **personality** and **emotions** should be dimensions of believability, and the related questions were asked by the overwhelming majority of the questionnaires reviewed in this study. **Visual properties (or appearance)**, **behavior**, **awareness**, and **social relationships** are also broadly recognized elements that contribute to the believability of virtual agents [Mat01, GPMJ13, LH15, BTS16].

The "illusion of life" is another critical aspect of believability that is commonly articulated in more practical forms of self-motivation, change, broadly capable, liveness, etc. [Mat01, Loy97, BTS16]. It is worth noting that the trending of such questions in believability questionnaires in recent years aligns with the rapid development of intelligent virtual agents. For the purpose of brevity, we organized the related questions into **intelligence** and **agency**.

Overall believability is a concise yet effective way to assess the perceived believability toward virtual agents, which is usually presented in questions in the form of rating the believability, human-likeness, or plausibility of the interaction [BZ09]. The related ques-

tions were asked in approximately half of the questionnaires reviewed in this study. In most cases, the related questions were asked in addition to questions that focused on single dimensions of believability, while some questionnaires revolved entirely around rating the overall believability of the virtual agents.

3.2. The Virtual Agent Believability Scale

After reviewing the believability questionnaires from the experiments in the past two decades, we proposed a set of items for designing the standardized scale for future studies. We organized the items into eight main dimensions of believability: visual properties, behavior, awareness, social relationships, intelligence, emotion, personality, and agency. We also included a few extra items on human's perceived overall believability to reflect the multidimensional nature of believability, and we presented them in the following list:

Visual properties (or appearance)

Q1. The visual design of the virtual agent caught my attention.
Q2. I think the virtual agent's appearance is aesthetically pleasing.
Q3. I think the virtual agent's visual design is realistic.

Behavior

Q4. The virtual agent's behavior drew my attention.
Q5. I felt the virtual agent's behavior was coherent and natural.
Q6. I think the virtual agent's behavior was easy to understand.
Q7. I felt the virtual agent's behavior was appropriate to the context.
Q8. I felt sometimes the virtual agent behaved inappropriately.

Awareness

Q9. I felt that the virtual agent perceived the environment around him/her/them.
Q10. I felt that the virtual agent reacted to the change in the environment.
Q11. I felt that the virtual agent was aware of my presence.
Q12. I felt that the virtual agent was aware of the presence of other virtual agents.
Q13. The virtual agent was unaware of its surroundings.

Social relationships

Q14. The virtual agent interacted socially with me.
Q15. I felt that the virtual agent was able to coordinate with me.
If there is more than one virtual agent.
Q16. "The virtual agent interacted socially with the other virtual agent(s).
Q17. I felt that the virtual agent was able to coordinate with the other virtual agent(s).

Intelligence

Q18. I felt that the virtual agent was able to make plans.

Q19. I felt that the virtual agent learned from past experiences.
Q20. I felt that the virtual agent seemed to have memory.

Emotion

Q21. I felt that the virtual agent was capable of having feelings.
Q22. I felt that the virtual agent expressed emotions.
Q23. I felt that the virtual agent's expressed emotions were easy to understand.
Q24. I felt that the virtual agent's expressed emotions were appropriate to the context.

Personality

Q25. I felt that the virtual agent had a personality.
Q26. I felt that the virtual agent was extraverted and enthusiastic.
Q27. I felt that the virtual agent was sympathetic and warm.
Q28. I felt that the virtual agent was dependable and self-disciplined.
Q29. I felt that the virtual agent was emotionally stable.
Q30. I felt that the virtual agent was open to new experiences.

Agency

Q31. I felt that the virtual agent seemed to have self-awareness.
Q32. The virtual agent took actions without inputs from others.
Q33. The virtual agent seemed to have its own goals.

Overall believability

Q34. I felt that the virtual agent was believable.
Q35. I felt that the virtual agent behaved like a real person.
Q36. I enjoy the interaction with the virtual agent.

Future researchers who adopt this scale in their experiments might use only part of the items or add extra items to address the questions of concern. For example, if an experiment involves only one virtual agent, items on the interactions between the agents would be unnecessary. However, in most cases, we recommend using the complete 36-item version for ease of drawing comparisons across experiments. We also recommend that researchers include the above items in a randomized order to minimize context effects.

We suggest that the experiments should use a 7-point Likert scale, ranging from strongly disagree (1), disagree (2), somewhat disagree (3), neither agree nor disagree (4), somewhat agree (5), agree (6), and strongly agree (7), since point scales were prevailing in previous studies we reviewed. They yield ordinal data and thus benefit the subsequent statistical analysis in research. In addition, compared to other forms of Likert scales, such as a 5-point scale, the 7-point Likert scale allows participants to differentiate their intuitive feelings or perceptions more precisely. Further, the items should be presented to the participants immediately after the condition or after each condition in within-subject studies.

If the experiment design needs to present the result as a single believability score instead of eight separate scores per dimension, we recommend that researchers calculate the scores as follows:

1. Visual properties = Q1 + Q2 + Q3
2. Behavior = Q4 + Q5 + Q6 + Q7 - Q8
3. Awareness = Q9 + Q10 + Q11 + Q12 - Q13
4. Social relationships = Q14 + Q15 + Q16 + Q17
5. Intelligence = Q18 + Q19 + Q20
6. Emotion = Q21 + Q22 + Q23 + Q24
7. Personality = Q25 + Q26 + Q27 + Q28 + Q29 + Q30
8. Agency = Q31 + Q32 + Q33

Otherwise, if the experimental design intends to present believability as a single score, we suggest that the total believability score should be calculated using the following formula, in which the weights denote the prevalence of questions from each dimension:

Total believability = (Visual properties/3 + (Behavior/5) $\times$ 2 + (Awareness/5) $\times$ 2 + Social relationships/4 + (Intelligence/3) $\times$ 3 + (Emotion/4) $\times$ 4 + (Personality/6) $\times$ 4 + (Agency/3) $\times$ 2 + Overall believability/3) / 20,

where Overall believability = Q30 + Q31 + Q32

The formula weights certain dimensions higher to reflect their significance (e.g., personality and emotion), which is in agreement with previous studies that have proposed metrics of believability [Loy97, AP09, PHN*09, GPMJ13]. The results of the scales are comparable both within and between subjects and across the experimental conditions if surveyed immediately after the experimental conditions. However, in cases where nonparametric tests need to be used in data analysis, the scores should be treated as non-continuous variables. We also suggest that consistency checks should be conducted on all responses in the experiments. In normal cases, the score of Q8 should be the inverse of Q7, and the score of Q9 should be the inverse of Q13.

3.3. Prevalence

In this section, we focus on the prevalence of the proposed items as well as their significance in assessing their respective dimensions of believability. As detailed in Table 2 in the supplementary material document, among the reviewed studies, 12 experiments utilized questionnaires with items beyond merely rating the experimental target's overall believability. Studies using only general rating questions have been omitted for organizational clarity. In light of the limited pool of virtual agent's believability questionnaires in directly relevant research, we have included questions from questionnaires that examine the believability of creatures and robots [PHN*09, BCR17, TWWV19]. We also included templates of believability questionnaires introduced by previous studies for a more comprehensive investigation.

In our proposed scale, we discarded questions that are not applicable to measuring virtual agents' believability. Examples include "Anna wanted to win the game" (1Q7), "I would prefer to rent/buy through a virtual character rather than a human real estate agent" (5Q12), and "The robot perceived the content of the movie clip correctly" (8Q1). Further, we excluded items of very low prevalence or significance, for example, "The character's recollection was nat-

ural" (5Q6), "I believe Axo is capable of experiencing embarrassment" (11Q6), and "The characters expressed emotions in anticipation of what could happen" (12Q6). The rest of the items were categorized based on relevant criteria, as illustrated in previous sections. For example, "I felt that the virtual agent has a personality" (Q25) is the combination of the questions "<X >has a personality" (3Q3), "The robot has a personality" (8Q4), "Suspect had a personality" (9Q1), and "The character has a personality" (10Q3).

From the reviewed studies, we identified visual properties (or appearance) as a dimension of believability. Although still an emerging concept, Bogdanovych et al. [BTS16] argued that a believable appearance contributes to the believability of virtual agents. To enhance the comprehensiveness of the proposed scale, we integrated the argument into two items: "I think the virtual agent's appearance is aesthetically pleasing" (Q2) and "I think the virtual agent's visual design is realistic" (Q3). For the special cases of Q26, Q27, Q28, Q29, and Q30, we used five items to assess the five dimensions of personality perceived by humans, which were adopted from the ten-item personality inventory (TIPI) to enable a brief yet accurate measurement of personality traits [GRSJ03]. For example, "I felt that the virtual agent is sympathetic and warm" (Q27) was the adaptation of the question items "Critical, quarrelsome" and "Sympathetic, warm" from the original personality inventory.

4. Discussion

In this paper, we propose a scale for measuring a virtual agent's believability. Our work is based on the questionnaires used in previous studies on believable virtual agents, characters, game creatures, and robots over the past two decades. Prior works from closely related fields, such as the personalities of virtual agents and their assessments, were also reviewed to ensure the comprehensiveness of the scale. The believability of virtual agents has been a popular field of research for years. Our review focuses on works from 1997 to the present. This work could be supplemented by future research to ensure its relevance and currentness.

In the review process, we identified eight main dimensions that contribute to the overall believability of virtual agents: visual properties, behavior, awareness, social relationships, intelligence, emotion, personality, and agency. This set of dimensions aligns with the prior literature on the metrics of believability. We categorized the questions used in previous research into eight dimensions and summarized them into a standardized scale while integrating question items from related research fields. We also included some additional items on overall perceived believability. The prevalence of the proposed items is presented in Table 3 in the supplementary materials document. Specifically, the prevalence is calculated by dividing the number of studies that used relevant questions by the total number of studies listed in Table 2 in the supplementary materials document.

This scale should be presented to the participants at the end of the study or the end of each experimental condition, depending on the experimental design. To calculate the final believability score, we suggested a simple formula. However, we recommend that future studies modify it based on their unique research questions.

Although our proposed scale awaits validation, our work is

grounded in previous research, and we hope for further validation from the research community. We recognize that certain items of our proposed scale may not be broadly used by directly relevant prior research. However, it is crucial to note that the existing body of literature on virtual agent believability is constrained by the lack of standardized measuring methods. The considerable variation in existing questionnaires calls for the proposal of a standardized scale to overcome the ongoing challenge of reproducibility faced by today's virtual reality experimentations [Grü23].

Acknowledgments

References

[ALA*05] AYDIN M. D., LEBLEBICI D. N., ARSLAN M., KILIC M., OKTEM M. K.: The impact of iq and eq on pre-eminent achievement in organizations: implications for the hiring decisions of hrm specialists. *The International Journal of Human Resource Management 16*, 5 (2005), 701–719. 2

[AP09] AFONSO N., PRADA R.: Agents that relate: Improving the social believability of non-player characters in role-playing games. In *Entertainment Computing-ICEC 2008: 7th International Conference, Pittsburgh, PA, USA, September 25-27, 2008. Proceedings 7* (2009), Springer, pp. 34–45. 2, 6

[APA13] AVRADINIS N., PANAYIOTOPOULOS T., ANASTASSAKIS G.: Behavior believability in virtual worlds: agents acting when they need to. *SpringerPlus 2*, 1 (2013), 1–11. 2

[AWFM23] ALJAMMAZ R., WARDRIP-FRUIN N., MATEAS M.: Towards an understanding of character believability. In *Proceedings of the 18th International Conference on the Foundations of Digital Games* (2023), pp. 1–9. 2

[BCR17] BARRETO N., CRAVEIRINHA R., ROQUE L.: Designing a creature believability scale for videogames. In *Entertainment Computing–ICEC 2017: 16th IFIP TC 14 International Conference, Tsukuba City, Japan, September 18-21, 2017, Proceedings 16* (2017), Springer, pp. 257–269. 2, 4, 6

[BKCZ09] BARTNECK C., KULIĆ D., CROFT E., ZOGHBI S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics 1* (2009), 71–81. 3

[BRDL17] BEVACQUA E., RICHARD R., DE LOOR P.: Believability and co-presence in human-virtual character interaction. *IEEE computer graphics and applications 37*, 4 (2017), 17–29. 2, 4

[BTS16] BOGDANOVYCH A., TRESCAK T., SIMOFF S.: What makes virtual agents believable? *Connection Science 28*, 1 (2016), 83–108. 1, 2, 4, 6

[BYA*12] BAILEY C., YOU J., ACTON G., RANKIN A., KATCHABAW M.: Believability through psychosocial behaviour: Creating bots that are more engaging and entertaining. *Believable Bots: Can Computers Play Like People?* (2012), 29–68. 2

[BZ09] BOSSE T., ZWANENBURG E.: There's always hope: Enhancing agent believability through expectation-based emotions. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (2009), IEEE, pp. 1–8. 4

[CAC*22] CURTIS C., ADALGEIRSSON S. O., CIURDAR H. S., MCDERMOTT P., VELÁSQUEZ J., KNOX W. B., MARTINEZ A., GAZTELUMENDI D., GOUSSIES N. A., LIU T., ET AL.: Toward believable acting for autonomous animated characters. In *Proceedings of the 15th ACM SIGGRAPH Conference on Motion, Interaction and Games* (2022), pp. 1–15. 1, 3, 4

[Col71] COLERIDGE S. T.: Biographia literaria, 1817. 1

[CVB*12] CAFARO A., VILHJÁLMSSON H. H., BICKMORE T., HEYLEN D., JÓHANNSDÓTTIR K. R., VALGARÐSSON G. S.: First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings 12* (2012), Springer, pp. 67–80. 2

[DCYSV19] DE CONINCK F., YUMAK Z., SANDINO G., VELTKAMP R.: Non-verbal behavior generation for virtual characters in group conversations. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2019), IEEE, pp. 41–418. 4

[DH00] DAWDA D., HART S. D.: Assessing emotional intelligence: Reliability and validity of the bar-on emotional quotient inventory (eq-i) in university students. *Personality and individual differences 28*, 4 (2000), 797–812. 2

[DJC*21] DICKINSON P., JONES A., CHRISTIAN W., WESTERSIDE A., MULLOY F., GERLING K., HICKS K., WILSON L., PARKE A.: Experiencing simulated confrontations in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–10. 4

[DNP11] DEMEURE V., NIEWIADOMSKI R., PELACHAUD C.: How is believability of a virtual agent related to warmth, competence, personification, and embodiment? *Presence 20*, 5 (2011), 431–448. 2, 4

[DOQ02] DAUTENHAHN K., OGDEN B., QUICK T.: From embodied to socially embedded agents–implications for interaction-aware robots. *Cognitive Systems Research 3*, 3 (2002), 397–428. 2

[DS03] DOBBYN C., STUART S.: The self as an embedded agent. *Minds and Machines 13* (2003), 187–201. 3

[DZA20] DAYLAMANI-ZAD D., ANGELIDES M. C.: Altruism and selfishness in believable game agents: Deep reinforcement learning in modified dictator games. *IEEE Transactions on Games 13*, 3 (2020), 229–238. 4

[EBB21] EVEN C., BOSSER A.-G., BUCHE C.: Assessing the believability of computer players in video games: A new protocol and computer tool. *Frontiers in Computer Science 3* (2021), 774763. 4

[EK17] ELSAYED S., KING D. J.: Affect and believability in game characters: A review of the use of affective computing in games. In *GAME-ON'2017, 18th annual Conference on Simulation and AI in Computer Games* (2017), EUROSIS, pp. 90–97. 4

[GGW07] GRAY H. M., GRAY K., WEGNER D. M.: Dimensions of mind perception. *science 315*, 5812 (2007), 619–619. 3

[GKK18] GAO B., KIM J.-I., KIM H.: Sensory and perceptual consistency for believable response in action feedback loop. In *Proceedings of Computer Graphics International 2018*. 2018, pp. 201–210. 1

[GPMJ13] GOMES P., PAIVA A., MARTINHO C., JHALA A.: Metrics for character believability in interactive narrative. In *Interactive Storytelling: 6th International Conference, ICIDS 2013, Istanbul, Turkey, November 6-9, 2013, Proceedings 6* (2013), Springer, pp. 223–228. 1, 2, 3, 4, 6

[GPS*20] GUIMARÃES M., PRADA R., SANTOS P. A., DIAS J., JHALA A., MASCARENHAS S.: The impact of virtual reality in the social presence of a virtual agent. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (2020), pp. 1–8. 2, 4

[GRSJ03] GOSLING S. D., RENTFROW P. J., SWANN JR W. B.: A very brief measure of the big-five personality domains. *Journal of Research in personality 37*, 6 (2003), 504–528. 6

[Grü23] GRÜBEL J.: The design, experiment, analyse, and reproduce principle for experimentation in virtual reality. *Human Spatial Perception, Cognition, and Behaviour in Extended Reality 4* (2023), 1069423. 7

[GTBH06] GORMAN B., THURAU C., BAUCKHAGE C., HUMPHRYS M.: Believability testing and bayesian imitation in interactive computer games. In *From Animals to Animats 9: 9th International Conference on Simulation of Adaptive Behavior, SAB 2006, Rome, Italy, September 25-29, 2006. Proceedings 9* (2006), Springer, pp. 655–666. 3, 4

[Het15] HETHERINGTON R.: The believability of hyper realistic characters in animated movies. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation*. 2015, pp. 1–9. 1, 4

[HHC18] HORTENSIUS R., HEKELE F., CROSS E. S.: The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems 10*, 4 (2018), 852–864. 2

[IBS11] IJAZ K., BOGDANOVYCH A., SIMO S.: Enhancing the believability of embodied conversational agents through environment-, self- and interaction-awareness. In *Conferences in Research and Practice in Information Technology Series* (2011). 2, 3, 4

[III*18] ISHIHARA M., ITO S., ISHII R., HARADA T., THAWONMAS R.: Monte-carlo tree search for implementation of dynamic difficulty adjustment fighting game ais having believable behaviors. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)* (2018), IEEE, pp. 1–8. 4

[LCM*23] LAM L., CHOI M., MUKANOVA M., HAUSER K., ZHAO F., MAYER R., MOUSAS C., ADAMO-VILLANI N.: Effects of body type and voice pitch on perceived audio-visual correspondence and believability of virtual characters. In *ACM Symposium on Applied Perception 2023* (New York, NY, USA, 2023), SAP '23, Association for Computing Machinery. 1

[LH15] LEE M. S., HEETER C.: Cognitive intervention and reconciliation-npc believability in single-player rpgs. *International Journal of Role-Playing*, 5 (2015), 47–65. 4

[Liv06] LIVINGSTONE D.: Turing's test and believable ai in games. *Computers in Entertainment (CIE) 4*, 1 (2006), 6–es. 2, 3

[Loy97] LOYALL A. B.: *Believable Agents: Building Interactive Personalities.* Tech. rep., CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE, 1997. 2, 3, 4, 6

[LS97] LESTER J. C., STONE B. A.: Increasing believability in animated pedagogical agents. In *Proceedings of the first international conference on Autonomous agents* (1997), pp. 16–21. 1, 3

[LSL17] LIU F., SHI Y., LIU Y.: Intelligence quotient and intelligence grade of artificial intelligence. *Annals of Data Science 4* (2017), 179–191. 2

[Mat01] MATEAS M.: An oz-centric review of interactive drama and believable agents. In *Artificial intelligence today: Recent trends and developments*. Springer, 2001, pp. 297–328. 1, 4

[MCJ97] MCCRAE R. R., COSTA JR P. T.: Personality trait structure as a human universal. *American psychologist 52*, 5 (1997), 509. 2

[MHC21] MOGRABI D. C., HUNTLEY J., CRITCHLEY H.: Self-awareness in dementia: a taxonomy of processes, overview of findings, and integrative framework. *Current Neurology and Neuroscience Reports 21*, 12 (2021), 69. 2

[Mor11] MORIN A.: Self-awareness part 1: Definition, measures, effects, functions, and antecedents. *Social and personality psychology compass 5*, 10 (2011), 807–823. 2

[MSM*11] MCRORIE M., SNEDDON I., MCKEOWN G., BEVACQUA E., DE SEVIN E., PELACHAUD C.: Evaluation of four designed virtual agent personalities. *IEEE Transactions on Affective Computing 3*, 3 (2011), 311–322. 2

[MU14] MITARAI K., UMEMURO H.: Amae and agency appraisal as japanese emotional behavior: influences on agent's believability. In *Proceedings of the second international conference on Human-agent interaction* (2014), pp. 189–192. 4

[PHN*09] POEL M., HEYLEN D., NIJHOLT A., MEULEMANS M., VAN BREEMEN A.: Gaze behaviour, believability, likability and the icat. *AI & society 24* (2009), 61–73. 3, 4, 6

[PLBW17] PAIVA A., LEITE I., BOUKRICHA H., WACHSMUTH I.: Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS) 7*, 3 (2017), 1–40. 2

[PP05] PRADA R., PAIVA A.: Believable groups of synthetic characters. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems* (2005), pp. 37–43. 2

[PTPL18] PACHECO C., TOKARCHUK L., PÉREZ-LIÉBANA D.: Studying believability assessment in racing games. In *Proceedings of the 13th international conference on the foundations of digital games* (2018), pp. 1–10. 4

[RB14] RICHARDS D., BRANSKY K.: Forgetmenot: What and how users expect intelligent virtual agents to recall and forget personal conversational content. *International Journal of Human-Computer Studies 72*, 5 (2014), 460–476. 3, 4

[RM17] RODRIGUES R., MARTINHO C.: Towards believable interactions between synthetic characters. In *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings 17* (2017), Springer, pp. 385–388. 4

[RPCP16] RAVENET B., PECUNE F., CHOLLET M., PELACHAUD C.: Emotion and attitude modeling for non-player characters. *Emotion in Games: Theory and Praxis* (2016), 139–154. 2

[RS06] RIEDL M. O., STERN A.: Believable agents and intelligent story adaptation for interactive storytelling. In *Technologies for Interactive Digital Storytelling and Entertainment: Third International Conference, TIDSE 2006, Darmstadt, Germany, December 4-6, 2006. Proceedings 3* (2006), Springer, pp. 1–12. 2

[SRM23] SILVA J., RODRIGUES R., MARTINHO C.: Believability, anticipation, and... timing improving believability through timing manipulation. In *Proceedings of the 18th International Conference on the Foundations of Digital Games* (2023), pp. 1–11. 2, 4

[TC] TJOKROSETIO A. N., CHOWANDA A.: Character believability enhancement using facial expression recognition to improve the players immersive experience. 2, 4

[TJT95] THOMAS F., JOHNSTON O., THOMAS F.: *The illusion of life: Disney animation*. Hyperion New York, 1995. 1, 3

[TWWV19] TSIOURTI C., WEISS A., WAC K., VINCZE M.: Multimodal integration of emotional signals from voice, body, and context: Effects of (in) congruence on emotion recognition and attitudes towards robots. *International Journal of Social Robotics 11* (2019), 555–573. 4, 6

[TYKS12] TOGELIUS J., YANNAKAKIS G. N., KARAKOVSKIY S., SHAKER N.: Assessing believability. *Believable Bots: Can Computers Play Like People?* (2012), 215–230. 1, 3

[vdHBR14] VAN DER HAM W. F., BROEKENS J., ROELOFSMA P. H.: The effect of dominance manipulation on the perception and believability of an emotional expression. In *Emotion Modeling: Towards Pragmatic Computational Models of Affective Processes*. Springer, 2014, pp. 101–114. 3, 4

[WJV13] WARPEFELT H., JOHANSSON M., VERHAGEN H.: Analyzing the believability of game character behavior using the game agent matrix. In *DiGRA Conference* (2013), pp. 1–11. 2

[ZZM19] ZELL E., ZIBREK K., MCDONNELL R.: Perception of virtual characters. In *ACM Siggraph 2019 Courses*. 2019, pp. 1–17. 1