Reducing Task Load with an Embodied Intelligent Virtual Assistant for Improved Performance in Collaborative Decision Making

Kangsoo Kim* University of Central Florida

Celso M. de Melo† US Army Research Laboratory

Nahal Norouzi‡ University of Central Florida

Gerd Bruder§ University of Central Florida

Gregory F. Welch¶ University of Central Florida

Abstract

Collaboration in a group has the potential to achieve more effective solutions for challenging problems, but collaboration per se is not an easy task, rather a stressful burden if the collaboration partners do not communicate well with each other. While Intelligent Virtual Assistants (IVAs), such as Amazon Alexa, are becoming part of our daily lives, there are increasing occurrences in which we collaborate with such IVAs for our daily tasks. Although IVAs can provide important support to users, the limited verbal interface in the current state of IVAs lacks the ability to provide effective non-verbal social cues, which is critical for improving collaborative performance and reducing task load.

In this paper, we investigate the effects of IVA embodiment on collaborative decision making. In a within-subjects study, participants performed a desert survival task in three conditions: (1) performing the task alone, (2) working with a disembodied voice assistant, and (3) working with an embodied assistant. Our results show that both assistant conditions led to higher performance over when performing the task alone, but interestingly the reported task load with the embodied assistant was significantly lower than with the disembodied voice assistant. We discuss the findings with implications for effective and efficient collaborations with IVAs while also emphasizing the increased social presence and richness of the embodied assistant.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Ubiquitous and mobile computing—Ubiquitous and mobile devices—Personal digital assistants; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

1 Introduction

In modern life, there are a number of problems that individuals cannot deal with or would likely end up with less effective solutions when working alone—including physical but also mental and cognitive tasks. As the concept of humans as "social by nature" developed from Aristotle's famous aphorism, we strategically collaborate with others to overcome problems [66]. Such collaboration in a group has the potential to achieve more effective solutions for challenging problems, but collaboration per se is not an easy task, rather a stressful burden that induces more mental/cognitive load while communicating with the collaboration partners [35]. If the team members do not communicate well with each other or are not supportive, the perceived task load will increase significantly because of the overhead caused by the collaboration. A large body of literature in social psychology and collaborative learning has investigated this phenomenon and identified the importance of reducing task load in collaborative situations while still emphasizing the improvement of task performance at the same time.

With the advent of advanced artificial intelligence (AI) and natural language processing (NLP), intelligent virtual assistants (IVAs) have experienced dramatic technological achievements in both research and commercial application fields [55]. While IVAs, such as Amazon Alexa, Google Assistant, and Apple Siri, are becoming part of our daily lives, there are increasing occurrences in which we collaborate with such IVAs for our daily tasks [54, 63]. IVAs do not just passively search and provide information requested by the users through simple verbal commands, but can also proactively understand the users' context and make suggestions as an intelligent collaborative entity. Although IVAs can provide such important support to users, the limited verbal interface in the current state of IVAs lacks the ability to provide effective non-verbal social cues, which is critical for improving collaborative performance and reducing workload. Augmented reality (AR) has the potential to overcome this challenge by providing a visual embodiment for the IVAs. A human-like visual representation could enrich the communicative channels that convey the assistant's status and intentions by presenting emotional expressions and gestures as non-verbal social behaviors [32].

Many findings in prior literature showed that task performance could be improved by the collaborative interaction with IVAs, but there was not enough attention on how the IVA interaction could actually cause more task load in the social collaboration context. This is particularly interesting with respect to the IVA embodiment, given that prior work on IVAs has been inconclusive about the impact of embodied agents on cognitive load [60].

Thus, in this paper, we investigate the effects of an IVA's visual embodiment on collaborative decision making, specifically focusing on the objective task performance and subjective perception of task load. In a within-subjects study, participants performed a desert survival task in three conditions: (1) performing the task alone, (2) working with a disembodied voice assistant, and (3) working with an embodied assistant. Our results show that both assistant conditions led to higher performance over when performing the task alone, but interestingly the reported task load with the embodied assistant was significantly lower than with the disembodied voice assistant. The findings are beneficial to design assistants for many daily tasks, but particularly for professional settings that require IVAs that are, simultaneously, helpful and able to minimize the user's cognitive load. We discuss the findings with implications for effective and efficient collaborations with IVAs while also emphasizing the increased social presence and richness of the embodied assistant.

In particular, we investigated the following research questions:

- **RQ1**: Does working with IVAs improve objective task performance in collaborative tasks?
- **RQ2**: Does working with IVAs increase subjective perception of task load in collaborative tasks?

*e-mail: kangsoo.kim@ucf.edu
†e-mail: celso.m.demelo.civ@mail.mil
‡e-mail: nahal.norouzi@knights.ucf.edu
§e-mail: bruder@ucf.edu
¶e-mail: welch@ucf.edu

- **RQ3**: Does the IVA's visual embodiment improve objective task performance in collaborative tasks?
- **RQ4**: Does the IVA's visual embodiment decrease subjective perception of task load in collaborative tasks?

This paper is structured as follows. Section 2 presents an overview of related work. Section 3 describes the human-subject study. The results are presented in Section 4 and discussed in Section 5. Section 6 concludes the paper.

2 RELATED WORK

In this section, we present previous work that studied collaborative problem solving and decision making, task load, and embodiment effects in the scope of interactions between humans and IVAs.

2.1 Collaborative Decision Making

Multiple studies have shown that people tend to treat and get influenced by IVAs in a similar way as they would with other humans [1, 58]. In the context of decision making and cooperation, understanding how people are influenced by IVAs and the behaviors that lead to these effects becomes more important.

Khooshabeh et al. studied an intelligent agent's capability to exert social influence through its sense of humor in a problem solving task (a lunar surviving scenario), finding that participants who perceived the seemingly humorous agents as not funny were not as likely to use its suggestions for the task while those who did were more influenced by it [28]. Kulms et al. also investigated sense of humor and its influence on an agent's perceived cooperativeness in the iterated prisoner's dilemma [39], but they found that the humorous agent was rated as less cooperative although it was perceived as more enjoyable. Khan and Sutcliffe found that participants were more inclined to comply with a visually more attractive agent compared to an unattractive one in a desert survival task [27]. Lucas et al. studied how the approach towards exerting social influence (i.e., factual vs. personal preference) can affect participants' decision making process in a desert survival task [47], and found that a facts-based approach was more influential.

In the context of health care, researchers also investigate the effects of utilizing IVAs as a collaborative or coaching partner to aid decision making [16,59] or promote healthier habits [6,9,61]. Byrne et al. utilized virtual pet agents to encourage adolescents towards healthier eating decisions, finding that participants who interacted with an agent capable of both positive and negative behavior were more likely to consume breakfast compared to interacting with no agent or one limited to positive emotions [9]. In a shared decision making task for prenatal testing, Zhang and Bickmore found that using an embodied virtual agent could increase users' knowledge and enhance their satisfaction with their decision [69].

2.2 Task Load in Collaboration

Task load, particularly mental/cognitive load, pertains to the effort the brain has to undergo to process, learn, and access information [26]. Several factors have been identified that influence cognitive load, but, broadly, it is possible to distinguish between intrinsic factors (e.g., task difficulty) and extrinsic factors (e.g., presentation of the task or collaboration process). The importance of reducing cognitive load for task performance has been studied across several domains [17]. In educational settings, multimedia technology has often been studied as an alternative to traditional textbook and classroom teaching, with results often showing improved learning with reduced cognitive load [3, 43, 60]. In life or death situations, such as medical [2] or military [19] settings, it is imperative that technology is able to provide critical information efficiently and seamlessly to professionals that are under immense cognitive, and possibly physical, load.

In collaboration, which inherently involves complex social activities, interaction and coordination of cognitive effort are required, and it is challenging to collaboratively converge on a group decision for a shared problem [37]. IVAs are an emerging technology that has seen the potential as an intelligent collaborative partner, which can facilitate daily tasks through intuitive and natural interaction with users [24,45]. The basic premise is that advances in natural language processing technology [42] enable more natural open-ended conversation with machines, which are able to then provide information or carry out users' instructions. Verbal communication is also socially richer than other forms of communication like text or email, as it can convey pragmatic and affective information [30]. Most current commercial systems, though, only have limited capabilities to convey this social richness to users through speech. This verbal communication can cause more cognitive load in the collaboration context due to the linguistic ambiguity. Overall, while easier access to information is expected to lead to improvements in task performance, the impact of these types of assistants on users' cognitive load is still not well understood.

2.3 Assistant Embodiment and Augmented Reality

IVA embodiment has been extensively researched in the past, with findings either advocating [5, 12] or questioning [22, 23] their necessity. These variation can be attributed to the wide range of agent appearances, behaviors, and domains utilized in previous research efforts [14, 38]. In support of a conversational agent's embodiment, an important factor is the utilization of both verbal and nonverbal communication channels, facilitating relational behaviors which in turn can help build and maintain relationships in order to initiate trust and smooth cooperation among collaborators [5].

Beun et al. described embodied IVAs as a helpful and motivating presence in learning tasks [4]. In their study, where participants were presented with stories either in a text format or through the embodied agents, participants showed significantly better recall in the agent conditions. In the context of a direction giving task, Hasegawa et al. varied the type of agent providing the directions (i.e., robot, embodied virtual agent, and GPS), and the agent embodiment proved to have positively influenced users' perception although it had no impact on their performance. To evaluate the level of users' trust in automated cars, Haeuslschmid et al. explored different visualization approaches (i.e., embodied agent, world in miniature, car indicators) to communicate the intent of the car. Their results suggest that competence is a more important factor compared to friendliness, a characteristic used for designing the agent, in developing trust in contexts where safety is critical [23]. Demeur et al. evaluated the impact of agent embodiment and emotion over the perceived social believability in the agent, and found that appropriate emotions conveyed through the agent's embodiment, particularly related to the sense of competence and warmth, could lead to higher believability [15].

As AR technology promises the pervasive ability for us to access contextually-relevant information through virtual entities [31, 68], research on embodied virtual agents in AR becomes more and more active. Kim et al. investigated the effects of an IVA's embodiment and locomotion behavior on the sense of social presence and confidence in the agent, and found that both factors positively impacted users' perception of the agent's ability to be aware of and influence the real world compared to the disembodied voice agent [32, 33]. They also showed the benefits of embodied IVA's environmental physical interaction through a multimodal interface [34]. Wang et al. also conducted a study investigating user preference for different types of embodied or disembodied IVAs in AR while performing a visual search task together [67], and showed that participants preferred the miniature embodied IVA. Beyond the agent systems, research on telepresence and collaboration through embodied virtual avatars in AR is also growing [25,57]. The previous findings and the variation in some of the results in terms of subjective perception and objective performance emphasize the importance of further research on collaboration with embodied IVAs in AR.

3 Experiment

In this section we present the experiment that we conducted to investigate the effects of different types of virtual assistants on task performance and cognitive load in problem solving. The experiment was approved by the Institutional Review Board of our university.

3.1 Participants

We recruited 37 participants from our local university population for our experiment, and 36 among them completed the entire experiment—one withdrew from the study for personal reasons. We further excluded two more participants due to a failure to record data; thus, we had 34 participants (25 male and 9 female, ages 18 to 33, $M = 21.9$, $SD = 4.1$) for the analysis. All of the participants had normal or corrected-to-normal vision—12 with glasses and 7 with contact lenses. On a 7-point scale (from 1=not familiar at all to 7=very familiar), the level of participant-reported familiarity with AR/MR technology was comparatively high ($M = 4.56$, $SD = 1.33$). All participants had fewer than ten times of AR head-mounted display (HMD) experiences, and it was the first experience for 13 of them. Participants were also asked about their frequency of using commercial virtual assistant systems, such as Amazon Alexa, Apple Siri, or Microsoft Cortana. Their responses varied from no use at all to frequent daily use: eight participants indicated multiple times per day, two indicated once a day, eight indicated once a couple of days, seven indicated once a week, three indicated once a month, and six indicated no use at all. Five participants had prior experience with the desert survival task or closely related tasks.

3.2 Materials

In this section, we describe the AR desert survival task environment that we developed for our experiment as an abstract problem solving challenge, and two different virtual assistant types incorporated in the survival task.

3.2.1 Desert Survival in AR

The desert survival task, which was developed by Lafferty [41], is one of the most widely used team building exercises, in which people have to prioritize fifteen items, such as a bottle of water or a jackknife, according to their importance for surviving in the desert. The task involves collaborative social skills and cognitive/mental load to make better decisions which made it a good candidate for many human-agent interaction studies that investigate the effects of socially engaging conversational agents [29]. The task performance can also be objectively evaluated with scores based on a solution sheet provided by Pond[1].

For our experiment, we developed an AR desert survival task environment, where participants had to place real image markers illustrating the fifteen survival items in the order of importance while experiencing AR visual and auditory stimuli (Figure 1). The image markers were attached on physical foam bases so that the participants could intuitively grab them and move them around. To initiate the task, participants first looked at the marker with a desert image and put it on a start placeholder, which was virtually displayed on the table through an optical see-through HMD, Microsoft HoloLens. Once the desert image marker was placed in the start placeholder, the instruction and state boards virtually appeared with fifteen item placeholders on the table, where participants could place the survival items in their chosen order. When the item was placed in one of the placeholders, the placeholder turned to blue with a clicking sound effect and a virtual image corresponding to the item image was shown on it. Participants could freely re-order the items and check the status of placed items via the state board while performing the task. After all the fifteen items were placed in the item placeholders, a finish placeholder was shown in AR next to the desert image marker, and the instruction guided the participants to put the desert marker on the finish placeholder to complete the task. Once the participants put the desert marker on the finish placeholder, the task was completed, showing a message that guided them to call the experimenter. Throughout the experiment, the performance score was continuously calculated by comparing the current state of item placements and the solution order. The size of the markers was $10\,\text{cm} \times 10\,\text{cm} \times 1\,\text{cm}$, and the PTC Vuforia Engine[2] was used for marker recognition.

[1] Alonzo W. Pond, M.A. was an survival expert who was a former chief of the Desert Branch of the Arctic, Desert, Tropic Information Center of the Air University at Maxwell Air Force Base.

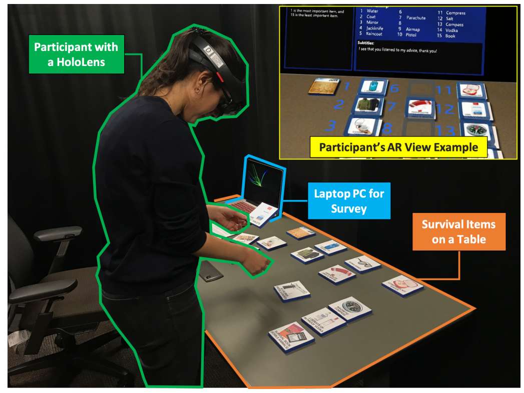


Figure 1: Physical setting in which a participant performs the desert survival task with fifteen image markers illustrating survival items. The participant experiences AR visual and auditory stimuli during the task.

3.2.2 Conversational Assistants

Here we describe two different virtual assistant types that we developed for the experiment: the *embodied* assistant and the non-embodied *voice* assistant:

- **Embodied** Assistant: A miniature of a 3D female character[3] was used for the embodied virtual assistant with visual appearance (see Figure 2). The character's blendshapes and LipSync[4] asset were used for lip movements during speech and facial expressions. The character was programmed to have a smiling and pleasant facial expression throughout the experiment, with a slight increase in the smile level during talking. Animations from Unity Standard Assets[5], Mixamo[6], and Inverse Kinematics[7] were used to augment the character with body gestures and idle behaviors. She could exhibit positive or negative expressions with body gestures using eleven different gesture animations, which include acknowledging nod, happy hand gesture, hard head nod, look down, talk with hand gestures, annoyed head shakes, sarcastic head nod, etc. She could also perform appreciation gestures, such as bowing down or putting one hand on her chest politely. The virtual character was superimposed over the real environment through a Microsoft HoloLens, which the participants wore, and could communicate with gestures and voice while walking on the table in front of the participants. The Unity Third Person Character Controller was used to generate the natural locomotion behavior so

[2] https://developer.vuforia.com/

[3] An in-house customized 3D human model were created based on Adobe Fuse (https://www.adobe.com/products/fuse.html) and FaceGen (https://facegen.com/) for the experiment.

[4] https://lipsync.rogodigital.com/

[5] https://assetstore.unity.com/packages/essentials/asset-packs/standard-assets-32351

[6] https://www.mixamo.com/

[7] https://docs.unity3d.com/Manual/InverseKinematics.html

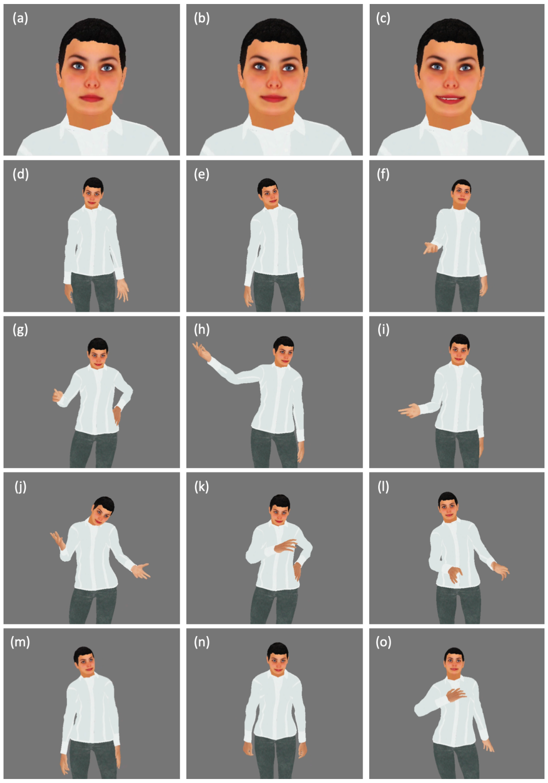


Figure 2: Examples of various facial expressions and body gestures of the embodied assistant: (a–c) different levels of pleasure: neutral, slight smiling, and more pleasant smiling, (d) idle standing animation, (e–m) different body gestures exhibiting the state of thinking and making suggestions while conveying the information about the survival items, and (n–o) gestures appreciating the participant's compliance when following a suggestion.

that the character could walk over the table while participants placed the survival items during the experiment. The realistic voice of the character was achieved by pre-recording her speech prompts to the participants using the Vocalware text-to-speech (TTS) API[8]. The speech prompts included overall task instructions, general acknowledgments, and the survival item suggestions. The assistant's speech was also displayed in text as subtitles on the state board.

- **Voice** Assistant: We disabled the visual embodiment of the embodied virtual assistant, so the voice assistant could not leverage any embodied human gestures or locomotion to convey social signals. In this way, participants could only hear the voice of the assistant through the HoloLens. This is similar to the current state of commercial assistant systems, e.g., Amazon Alexa, which communicate with users mainly through verbal interaction.

3.2.3 Physical Setup

Participants performed the desert survival task in an isolated space with black curtains around, so that they could concentrate on the task and the assistants' verbal and/or gestural behaviors. The space had a table, on which the participants placed the survival items, and a laptop PC that they used to answer the questionnaires (see Figure 1).

3.3 Methods

To examine the effects of the assistant's presence and embodiment, a within-subjects design was used for our experiment with three different conditions (see Figure 3):

- **Control** (No Assistant): As a baseline condition, participants performed the desert survival task by themselves without any assistants. In this condition, participants only saw the virtual boxes to indicate the places to locate the survival items and virtual boards describing the current state of the task and instructions.

- **Voice** Assistant: The virtual assistant did not have any visual appearance in the HoloLens, but participants could hear her voice while performing the survival task as described in the *Conversational Assistants* section above.

- **Embodied** Assistant: The miniature virtual human character described in the *Conversational Assistants* section was augmented with a visual embodiment, i.e., a human-like body, on the table. During the desert survival task, the assistant was trying to help the participants make better decisions for their survival by making suggestions that could potentially improve the task score. While making suggestions, the assistant walked toward the suggested item and presented different body gestures to convey the information about the item more effectively with spatial and emotional gestures and expressions.

The order of the conditions was counter-balanced with 36 participants who completed the experiment (see Section 3.1). In this way, we reduced the carryover effects between the three conditions, while having the participants directly compare the experiences based on their individual baseline.

3.3.1 Interaction Scenario

During the experiment, participants performed the desert survival task using the AR environment in collaboration with the different assistants according to the study condition. While participants performed the task alone in the Control condition without any virtual assistant, the assistants in the Voice and the Embodied conditions were trying to help the participants make better decisions during the task by providing suggestions that could potentially improve the task score. The system recognized where the items were currently located during the task and calculated the current survival score continuously. In this way, the assistants could determine the item that she would suggest to move, such that the participants could make the largest improvement in the survival score if they followed the suggestion accordingly. There were both positive and negative information prompts prepared for each survival item, so the assistant suggested to move a given placed item either up or down from its current place while providing the positive or negative information about the item. For example. the positive suggestion for the flashlight was "The flashlight could be a quick and reliable night signaling device. Why don't you move it up a bit?" and the negative suggestion was "I think the flashlight battery will be gone very quickly, so it might not be as useful as you expect. Why don't you move it down a bit?" There were three different prompt variations for both moving up and down suggestions, e.g., "I think it's ok to move it up/down," and "I guess you can move it up/down more." The assistant could also make stronger suggestions expressing that the item position should be adjusted a lot. For example, "I think you should move it up/down a lot," "Why don't you move it up/down a lot in the ranking?" and "I'm sure you can move it up/down quite a lot." The assistants could make the same suggestions repeatedly if the item was still the best option to improve the task score; however, if there was nothing to change for the score, no suggestion was provided. When suggesting items in the Embodied condition, the assistant walked towards and stood by the suggested item, and provided body gestures with more pleasant facial expressions while talking about the item. In this way, the embodied assistant could provide richer social cues to convey the information to the participants more effectively and convincingly. Participants received up to ten suggestions from the assistant by the

[8]https://www.vocalware.com

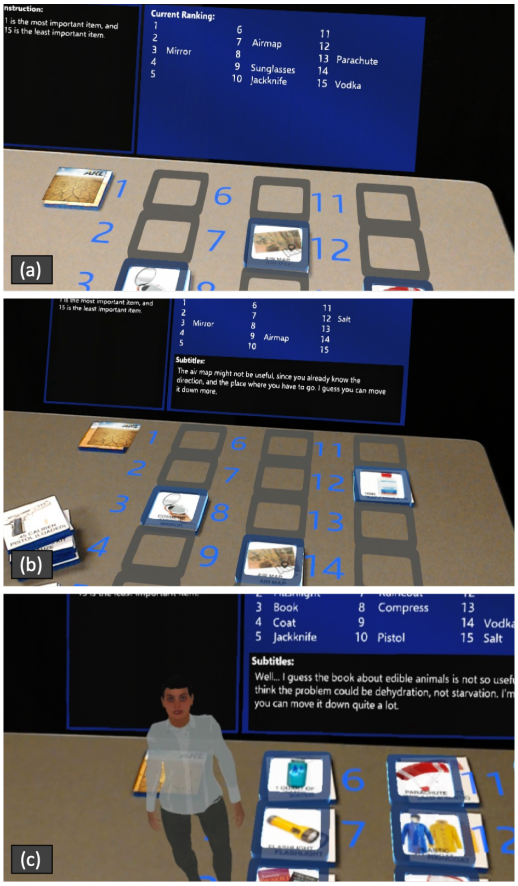


Figure 3: Experimental conditions: (a) Control: participants perform the desert survival task without any assistants, (b) Voice assistant: participants hear the conversational assistant's speech but cannot see her embodied appearance, while the assistant tried to help them out by making suggestions, (c) Embodied: participants can both see and hear the virtual assistant.

completion of the task. It is important to know that we guided the participants to decide whether they would follow the assistant's suggestions or not; thus, if they wanted to, they could keep the current item priority as it was. Once the participants actually followed the suggestions, the assistants performed appreciation prompts, such as, "Thank you for listening to my suggestion," which could encourage more compliance by participants for follow-up suggestions. The assistant also gave task instructions and general acknowledging comments, which included some variations of simple assuring comments, such as "Okay," "Good," or "You are doing great so far." These general comments could also encourage the participants to perform the task more actively and make the assistant's behavior appear more natural and interactive. The embodied assistant had an idle animation, which had slight body movements, and random eye-blinking to avoid an unnaturally static pose.

3.3.2 Procedure

Once participants arrived, they were guided to our laboratory space by the experimenter. They were asked to sit down in a room with a table and a laptop PC for answering questionnaires, and were provided with the consent form. Once they agreed to participate in the experiment, they donned a HoloLens and went through the calibration procedure on the HoloLens to set their interpupilary distance. Afterward, participants had a practice session to learn how to use our marker-based interface by placing five animal markers. In this practice phase, they were asked to place the five animal markers in their preferred order on the table while experiencing AR visual feedback (see Section 3.2.1). The experimenter was present next to the participants to answer any questions that they might have during the practice phase, while explaining the way to place and re-order the items. Once they felt comfortable with the marker-based interface, the experimenter described their actual task, the desert survival task, and the goal to prioritize the fifteen items for their survival in a desert. In the description, participants were told that they were going to take part in the same task three times with some variations. Then, the first session started with one of the experimental conditions: either the Control, the Voice, or the Embodied condition as described in Section 3.3. After completing the task, the participants were guided to complete several questionnaires measuring their perception of the experience in the desert survival task with or without assistant. When they were done answering the questionnaires, the experimenter guided them to repeat the same task in the next condition. Once the participants completed all three conditions, they answered further demographics and prior experience questionnaires, assessing their familiarity with AR and virtual assistant technology and experience. At the end, the participants were provided with a monetary compensation. The entire experiment took about an hour for each participant.

3.3.3 Hypotheses

We established two hypotheses for each of task performance and task load in collaborative decision making with virtual assistants, and one hypothesis for each of social presence and social richness perception, based on previous literature supporting the importance of assistant embodiment:

- **H1a**: Participants' task performance score in the Voice condition will be higher than in the Control condition (i.e., Control $<$ Voice).
- **H1b**: Participants' task performance score in the Embodied condition will be even higher than in the Voice condition (i.e., Voice $<$ Embodied).
- **H2a**: Participants' evaluation of task load in the Embodied condition will be lower than in the Voice condition (i.e., Embodied $<$ Voice).
- **H2b**: Participants' evaluation of task load in the Control condition will be even lower than in the Embodied condition (i.e., Control $<$ Embodied).
- **H3**: Participants' sense of social presence in the Embodied condition will be higher than in the Voice condition (i.e., Voice $<$ Embodied).
- **H4**: Participants' sense of social richness in the Embodied condition will be higher than in the Voice condition (i.e., Voice $<$ Embodied).

3.3.4 Measures

In this section, we describe our measures used to assess the influence of presence and type of virtual assistants in the study.

- **Task Performance**: The desert survival task has a solution sheet provided by a survival expert (see Section 3.2.1). We calculated the sum of absolute differences (SAD) between the order of the items provided by participants and the solution sheet, and negated the value for a more intuitive representative score. In this way, the best score is zero, which means all the items the participants placed are identical to the solution sheet. The task performance gets worse as the score moves further along the negative scale. The AR desert survival task environment automatically calculated and stored the final score at the end of each task.

- **Task Load**: The NASA Task Load Index (NASA-TLX) questionnaire [21] was used to assess the task load. The NASA-TLX is a standard measure in human factors and ergonomics evaluations, consisting of six questions, each corresponding to one dimension of the perceived workload. For each dimension (mental demand, physical demand, temporal demand, performance, effort, and frustration), the participants provide a score on a scale, from "Very Low" to "Very High," consisting of 21 tick marks effectively identifying 5% delimitations on a scale of 0% to 100%. Participants then provide weights for each of the six dimensions via a series of binary choices to assess which dimensions were most important for the task; these weights are then factored into the final score by multiplying them with the dimension scores.

- **Social Presence**: We adopted the social presence sub-scale from the Temple Presence Inventory (TPI) questionnaire [44] and slightly modified it to assess participants' sense of togetherness in the same space with the assistant, and the quality of the communication/interaction between them. The scale consists of seven questions on a 7-point scale from 1=not at all to 7=very much. We used this questionnaire only for a subjective comparison of the assistant conditions, i.e., the Voice and the Embodied conditions.

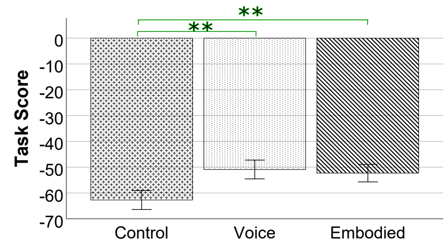
- **Social Richness**: We adopted the social richness sub-scale from the TPI questionnaire [44] to assess the extent to which the assistant is perceived as immediate, emotional, responsive, lively, personal, sensitive, and sociable. All the items for social richness are 7-point semantic differential scales, e.g., 1=remote and 7=immediate. We also used this questionnaire only for the assistant conditions, i.e., the Voice and the Embodied conditions.

4 RESULTS

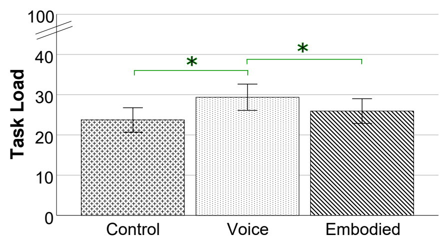
In this section, we report the results for our experimental measures. We used repeated measures ANOVAs and estimated marginal means for the post-hoc pairwise comparisons at the 5% significance level with Bonferroni correction in line with the ongoing discussion in the statistics literature suggesting that these parametric tests can be a valid and informative method for the analysis of combined experimental questionnaire scales as described above [36, 40, 53]. For cases where sphericity was not assumed through Mauchly's test and the Greenhouse-Geisser epsilon was larger than .75, we used Huynh-Feldt corrections for sphericity. Shapiro-Wilk tests and Q-Q plots were used to test for normality. Effect sizes are reported for the significant effects. For the measures specific for the perception of the assistants, we used paired-samples t-tests to compare the Voice and the Embodied conditions.

4.1 Task Performance

Task performance was evaluated by comparing the scores for the desert survival among the conditions. As we described in Section 3.3.4, zero was the best score, meaning that the participant's answer was exactly the same as the survival expert's answer, and the score was negated to make the larger score (closer to zero) represent the higher (better) score. The results are shown in Figure 4 (a). A repeated measures ANOVA showed that there was a statistically significant difference in the survival scores among the three conditions, $F(2,66)=10.17$, $\boldsymbol{p<.001}$, $\eta^2=.24$ (a large effect size). Post-hoc tests revealed that the scores in the Voice condition ($M=-50.88$, $SD=21.26$) were significantly higher than the scores in the Control condition ($M=-62.71$, $SD=21.38$), $\boldsymbol{p=.002}$, and the scores in the Embodied condition ($M=-52.29$, $SD=19.81$) were also significantly higher than the scores in the Control condition, $\boldsymbol{p=.002}$. There was no significant difference in the performance scores between the Voice and the Embodied conditions ($p=.99$). We also compared the participants' compliance rate for the assistants' suggestions between the Voice and the Embodied conditions, but did not find any statistical significance ($p=.95$). This indicates that both voice and embodied assistants were helpful for the participants to make better decisions in the scope of the desert survival task.



(a) Task Performance



(b) Task Load

Figure 4: Bar charts showing the results for (a) the task performance score, i.e., the desert survival score (higher is better), and (b) the NASA-TLX task load score (lower is better). For the task performance, the Voice and Embodied conditions showed better performance compared to the Control condition. For the task load, the Voice condition caused more task load than the Control and the Embodied conditions. The error bars show ±1 standard error of the mean. Statistical significance: ** (p<0.01), * (p<0.05).

4.2 Task Load

Following the established method, the NASA-TLX scores were calculated by summing the weighted sub-dimension scores [21]. The overall task load results are shown in Figure 4 (b) and the dimensions are shown in Figure 5. We found a statistically significant difference with a repeated measures ANOVA among the three conditions, $F(1.63,53.83)=5.24$, $\boldsymbol{p=.012}$, $\eta^2=.14$ (a medium to large effect size). Post-hoc tests revealed significant differences between the Control condition ($M=23.73$, $SD=17.69$) and the Voice condition ($M=29.37$, $SD=19.01$), $\boldsymbol{p=.015}$ and between the Voice condition and the Embodied condition ($M=25.94$, $SD=17.86$), $\boldsymbol{p=.026}$, but not between the Control and the Embodied conditions ($p=.86$). This indicates that the participants experienced a lower level of task load when they performed the task alone or with the embodied assistant compared to with the voice assistant.

4.3 Social Presence

Paired-samples t-tests revealed a statistically significant difference between the Voice ($M=3.59$, $SD=1.15$) and the Embodied ($M=4.60$, $SD=1.10$) conditions on the participants' sense of social presence with the virtual assistant, $t(33)=-4.57$, $\boldsymbol{p<.001}$, $d=0.79$ (a medium to large effect size). This indicates that the participants had a higher sense of social presence with the embodied assistant. The results are shown in Figure 6 (a).

4.4 Social Richness

Paired-samples t-tests revealed a significant main effect of the assistant embodiment on the perceived social richness of the assistant,

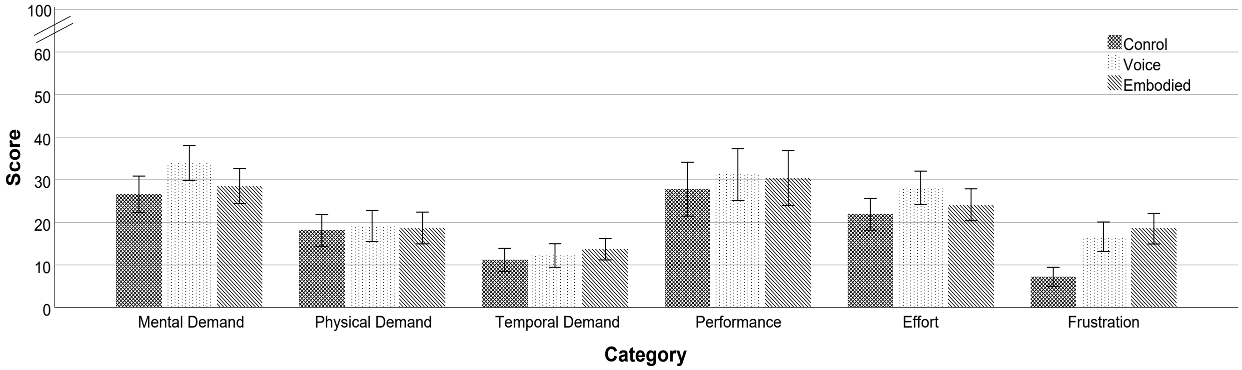
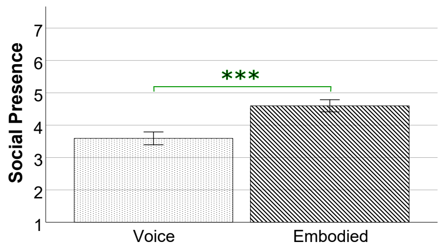
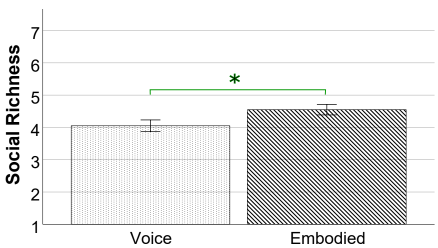


Figure 5: Bar charts showing the results for the individual dimensions of the NASA-TLX (task load) questionnaire: mental demand, physical demand, temporal demand, performance, effort, and frustration. The error bars show ±1 standard error of the mean.



(a) Social Presence



(b) Social Richness

Figure 6: Bar charts showing the results for (a) social presence and (b) social richness. In both cases, higher is better. For these measures, we only compared the two assistant conditions, and for both measures, the Embodied condition provided more social presence and richness than the Voice condition. The error bars show ±1 standard error of the mean. Statistical significance: *** (p<0.001), ** (p<0.01), * (p<0.05).

$t(33) = -2.57$, $\boldsymbol{p = .015}$, $d = 0.44$ (a small to medium effect size). The participants reported a higher perception of social richness in the Embodied condition ($M = 4.55$, $SD = 0.95$) compared to the Voice condition ($M = 4.05$, $SD = 1.06$), which indicates that the embodied assistant provided a richer and more immediate social perception to participants. The results are shown in Figure 6 (b).

5 Discussion

In this section, we summarize the main findings and discuss implications for the use of embodied virtual assistants in human-agent collaborative tasks.

5.1 Embodied Virtual Assistants Improve Task Performance and Reduce Task Load in Collaborative Problem Solving

Overall, our results show that participants achieved better performance with both disembodied voice and embodied assistants compared to performing the task alone, while the embodied assistant additionally helped them maintain a lower task load than experienced with the disembodied assistant.

For the task performance, in line with our Hypothesis **H1a**, we found that participants in the Voice condition performed the desert survival task with better scores than the Control condition. We initially expected the performance in the Embodied condition would even exceed the Voice condition, but in contrast to our Hypothesis **H1b**, we found that the performance in the Embodied condition was at a similar level as in the Voice condition. The participants' informal comments generally support the notion that the information provided by the assistants was perceived positively, and that is was perceived as more dominant for the decision-making task than the appearance of the assistants:

> P4: *"the assistant was very helpful, giving critical information in such a stressful situation if it happens in real world."* P7: *"The interaction of the assistant was overall beneficial, as it brought up many things I wouldn't have thought of."* P12: *"The information provided by the assistant were great, it helped me prioritize items better."*

Our experimental results confirm that both types of voice-only and embodied assistants can improve performance in an abstract collaborative problem solving task, when compared to performing the task without the support provided by an assistant. Our results are, in particular, interesting for commercial voice-only assistants, such as Amazon Alexa, given that they have received considerable attention from the industry. Our results reinforce that, even though there may be a cost, in practical terms, it is possible to improve performance and productivity through the careful application of assistant technology.

For the task load, in line with our Hypothesis **H2a**, we found that participant-reported task load in the Embodied condition was lower than in the Voice condition, and even maintained a similar level as in the Control condition, which is different from what we expected in our Hypothesis **H2b**. The results reveal that, even though both assistants caused an increase in performance, the embodied assistant accomplished this with minimal impact on cognitive load, when compared to the voice-only assistant. This is particularly interesting given that prior work on pedagogical agents has been inconclusive about the impact of embodied agents on cognitive load [60]: In some cases, post-tests were easier after interaction with an embodied agent [51]; in other cases, embodied agents led to increased mental effort for learners, even though there was no impact on performance [10]. According to Mayer's multimedia learning theory [48], there are two fundamental channels (auditory and visual) and optimal learning occurs when information is optimized across the channels (e.g., embodied agents will not produce an effect if they are redundant or irrelevant to the task [11]). In our case, though, the embodied assistant was serving clear functions above and beyond the voice-only assistant: through facial expressions, it smiled when making suggestions; through its virtual body, it moved and pointed to the target of the suggestions; and, generally, through subtle cues (e.g., idle motion or blinking), the assistant conveyed a human-like presence in the task. The experimental results confirm that these kinds of nonverbal cues have meaningful impact in lowering cognitive load for users. Participants' comments, such as the one below, support the notion of a benefit of visual embodiment for helping participants feel more comfortable in collaborative situations with virtual assistants:

> P28: *"I like that the assistant is naturally in front of you and given at the same time as I worked rather than pausing just to listen to what she had to say."*

5.2 Embodied Virtual Assistants Increase Social Presence and Richness

Our results showed that participants experienced higher social presence with the Embodied assistant than the Voice assistant, supporting our Hypothesis **H3**. Social presence relates to the ability of a communication medium to convey the sense that the user is immersed in the communication space and engaging in social interaction just as if it were face-to-face interaction [46, 62]. Research indicates that immersive technology has the potential to provide an increased sense of social presence, when compared to other media (e.g., phone or desktop) [7]. Our results support the notion that the embodiment of the virtual assistant in AR afforded an increased immersion and social presence during the interaction with the assistant, which may have contributed to the observed low impact on cognitive load.

Our results further indicate that participants perceived an increased social richness with the Embodied assistant than the Voice assistant, in line with our Hypothesis **H4**. This suggests that they were more likely to treat interaction with the embodied assistant in a social manner, as if they were interacting with another human. This is in line with prior research indicating that increased human-like cues [58] and immersion [7] can lead users to treat human-agent interaction like human-human interaction [49, 58], which can lead to positive effects in terms of engagement, motivation, and interest [3, 50, 60, 65]. The social richness of the experience with the embodied assistant, thus, may have additionally played a role in reducing the participants' cognitive load while performing the task.

Participants' informal comments about the Embodied condition are also in line with the notion that they perceived the assistant more like a real collaboration partner, and emphasize the benefits of the embodiment for increased social presence and richness:

> P24: *"Two heads are always better than one."* P26: *"It felt like I was communicating with another more knowledgeable person."*

5.3 Implications and Limitations

Task/cognitive load can actually be reduced by sharing additional cues in many collaborative situations. In that sense, simple primitive AR annotations, such as virtual arrows or other types of highlight indicators, could also be helpful in our study setting. This is, in fact, an interesting line of future inquiry. Embodied assistants, however, have the unique capability to engage users multimodally—like humans do [20]—and complement the information conveyed through speech with appropriate gestures and emotion beyond simple annotations. Our expectation, thus, is that embodied assistants provide a unique advantage. Given increasing evidence of the important role of nonverbal and emotional expression in social interaction [8, 13, 18, 52, 64], developers and designers cannot afford to ignore the value of embodiment for assistant technology. Our results support that, by using nonverbal cues judiciously, we can reduce the users' cognitive load without reducing performance. We should also note that there is additional value with the assistant embodiment in social context, e.g., the improved social presence and richness.

However, the current work also has limitations that introduce opportunities for future work. First, our current speech synthesizer—like most commercial systems—has limited expressive ability, which some of participants also commented. As speech technology improves, it will become possible to increase the bandwidth of multimodal expression [20]. Optimized multimodal expression can, then, lead to optimized transfer of information, learning, and performance [48]. Second, we used a first-generation Microsoft Hololens in our experiment, and, in practice, most participants still complained about the weight and bulkiness of the device. As AR head-mounted displays become better (e.g., lighter and supporting wider fields-of-view), we can expect increased immersion and impact of embodied assistants. Third, the current within-subjects design with a relatively small sample size could influence the participants' performance and perception. Future work should complement the present work with between-subject designs. Still, when we compared the participants' first trials as between-subjects comparisons, we found promising trends corresponding to our present results although not all the measures showed statistical significances, which encourages us to consider a further investigation in a between-subjects design with a large sample size. Finally, the current prototype only implemented basic AI (e.g., to determine optimal suggestions and whether the participant followed suggestions), but it is possible to embed more intelligence and autonomy into embodied assistants. In the field of human-agent interaction research, enhanced dialog models have been developed and researched to overcome such limitations [56]. The fast pace of development in technology across all these fronts and experimental research such as the one presented here that clarifies how best to use this technology, introduce a unique opportunity to create assistant technology that is immersive, feels like social interaction, is engaging and, most importantly, can promote optimal performance in our personal, social, and professional lives.

6 Conclusion

In this paper, we presented a human-subject study, in which we investigated the effects of the AR visual embodiment of an IVA on collaborative decision making in three conditions: performing the task alone, collaborating with a disembodied voice assistant, and collaborating with an embodied assistant.

Our results show that both the embodied and disembodied assistants led to significantly higher task performance compared to performing the task alone, while the embodied assistant further helped users maintain a significantly lower task load than with the disembodied assistant. We discussed the findings with respect to effective and efficient collaboration with IVAs, while also emphasizing the increased social presence and richness with the embodiment.

For future work, we believe that it would be interesting to investigate the influence of a longer familiarization period with the AR assistants on their effectiveness as a collaboration partner, as well as the inclusion of a natural language interface for an increased bi-directional interaction among the real and virtual interlocutors.

Acknowledgments

This material includes work supported in part by the National Science Foundation under Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the Office of Naval Research under Award Number N00014-17-1-2927 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting institutions. This research was partially supported by the US Army. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

References

[1] J. Appel, A. von der Pütten, N. C. Krämer, and J. Gratch. Does humanity matter? analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*, 2012:13, 2012.

[2] N. Archip, O. Clatz, S. Whalen, D. Kacher, A. Fedorov, A. Kot, and et al. Non-rigid alignment of pre-operative mri, fmri, and dt-mri with intraoperative mri for enhanced visualization and navigation in image-guided neurosurgery. *Neuroimage*, 35:609–624, 2007.

[3] R. Atkinson. Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94:416–427, 2002.

[4] R.-J. Beun, E. De Vos, and C. Witteman. Embodied conversational agents: effects on memory performance and anthropomorphisation. In *International Workshop on Intelligent Virtual Agents*, pages 315–319. Springer, 2003.

[5] T. Bickmore and J. Cassell. Relational agents: a model and implementation of building user trust. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 396–403, 2001.

[6] T. W. Bickmore, D. Mauer, and T. Brown. Context awareness in a handheld exercise agent. *Pervasive and Mobile Computing*, 5(3):226–235, 2009.

[7] J. Blascovich, J. Loomis, A. Beall, K. Swinth, C. Hoyt, and J. Bailenson. Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, 13(2):103–124, 2002.

[8] R. Boone and R. Buck. Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, 27:163–182, 2003.

[9] S. Byrne, G. Gay, J. Pollack, A. Gonzales, D. Retelny, T. Lee, and B. Wansink. Caring for mobile phone-based virtual pets can influence youth eating behaviors. *Journal of Children and Media*, 6(1):83–99, 2012.

[10] S. Choi and R. Clark. Cognitive and affective benefits of an animated pedagogical agent for learning english as a second language. *Journal of Educational Computing Research*, 24:441–466, 2006.

[11] S. Craig, B. Gholson, and D. Driscoll. Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94:428–434, 2002.

[12] C. de Melo, P. Carnevale, and J. Gratch. The Impact of Emotion Displays in Embodied Agents on Emergence of Cooperation with People. *Presence: Teleoperators & Virtual Environments*, 20(5):449–465, 2011.

[13] C. de Melo, P. Carnevale, S. Read, and J. Gratch. Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, 106:73–88, 2005.

[14] D. Dehn and S. Van Mulken. The Impact of Animated Interface Agents: A Review of Empirical Research. *International Journal of Human-Computer Studies*, 52:1–22, 2000.

[15] V. Demeure, R. Niewiadomski, and C. Pelachaud. How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment? *Presence: Teleoperators & Virtual Environments*, 20(5):431–448, 2011.

[16] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, et al. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1061–1068, 2014.

[17] A. Dey, M. Billinghurst, R. Lindeman, and J. Swan. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5, 2018.

[18] N. Frijda and B. Mesquita. The social roles and functions of emotions. In S. Kitayama and H. Markus, editors, *Emotion and culture: Empirical studies of mutual influence*. American Psychological Association, Washington, DC, 1994.

[19] J. Gillis. Warfighter trust in autonomy. *DSIAC Journal*, 4:416–427, 2017.

[20] J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan, and N. Badler. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, 17:54–63, 2002.

[21] S. Hart. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society*, pages 904–908, 2006.

[22] D. Hasegawa, J. Cassell, and K. Araki. The role of embodiment and perspective in direction-giving systems. In *AAAI Fall Symposium Series*, 2010.

[23] R. Häuslschmid, M. von Buelow, B. Pfleging, and A. Butz. Supporting trust in autonomous driving. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, pages 319–329, 2017.

[24] S. Henderson and S. Feiner. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37:81–88, 2018.

[25] D. Jo, K.-H. Kim, and G. J. Kim. SpaceTime: adaptive control of the teleported avatar for improved AR tele-conference experience. *Computer Animation and Virtual Worlds*, 26:259–269, 2015.

[26] S. Kalyuga. Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, 39:1–19, 2011.

[27] R. F. Khan and A. Sutcliffe. Attractive agents are more persuasive. *International Journal of Human-Computer Interaction*, 30(2):142–150, 2014.

[28] P. Khooshabeh, C. McCall, S. Gandhe, J. Gratch, and J. Blascovich. Does it matter if a computer jokes. In *ACM CHI Extended Abstracts on Human Factors in Computing Systems*, pages 77–86, 2011.

[29] C. Kidd and C. Breazeal. Effect of a Robot on User Perceptions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3559–3564, 2004.

[30] S. Kiesler, J. Siegel, and T. McGuire. Social psychological aspects of computer-mediated communication. *American Psychologist*, 39:1123–1134, 1984.

[31] K. Kim, M. Billinghurst, G. Bruder, H. Duh, and G. Welch. Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008-2017). *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2947–2962, 2018.

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

[33] K. Kim, N. Norouzi, T. Losekamp, G. Bruder, M. Anderson, and G. Welch. Effects of Patient Care Assistant Embodiment and Computer Mediation on User Experience. In *Proceedings of IEEE International Conference on Artificial Intelligence and Virtual Reality*, pages 17–24, 2019.

[34] K. Kim, R. Schubert, J. Hochreiter, G. Bruder, and G. Welch. Blowing in the wind: Increasing social presence with a virtual human via environmental airflow interaction in mixed reality. *Computers & Graphics*, 83:23–32, Oct 2019.

[35] P. Kirschner, J. Sweller, F. Kirschner, and J. Zambrano. From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning*, 13(2):213–233, 2018.

[36] T. R. Knapp. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing Research*, 39(2):121–123, 1990.

[37] G. L. Kolfschoten and F. M. Brazier. Cognitive Load in Collaboration: Convergence. *Group Decision and Negotiation*, 22(5):975–996, 2013.

[38] N. C. Krämer, B. Tietz, and G. Bente. Effects of embodied interface agents and their gestural activity. In *International Workshop on Intelligent Virtual Agents*, pages 292–300. Springer, 2003.

[39] P. Kulms, S. Kopp, and N. Krämer. Let's be serious and have a laugh: Can humor support cooperation with a virtual agent? In *Proceedings of the International Conference on Intelligent Virtual Agents*, pages 250–259, 2014.

[40] W. Kuzon Jr, M. Urbanchek, and S. McCabe. The seven deadly sins of statistical analysis. *Annals of Plastic Surgery*, 37(3):265–272, 1996.

[41] J. C. Lafferty and P. M. Eady. *The Desert Survival Problem*. Experimental Learning Methods, Plymouth, Michigan, 1974.

[42] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[43] J. Lester, S. Converse, S. Kahler, S. Barlow, B. Stone, and R. Bhogal. The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 359–366, 1997.

[44] M. Lombard, T. B. Ditton, and L. Weinstein. Measuring presence: The temple presence inventory. In *International Workshop on Presence*, pages 1–15, 2009.

[45] G. López, L. Quesada, and L. Guerrero. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In I. Nunes, editor, *Advances in Human Factors and Systems Interaction (AHFE)*. Springer, Cham, 2018.

[46] P. Lowenthal. *Social Presence*, pages 202–211. IGI Global, 2010.

[47] G. M. Lucas, J. Lehr, N. Krämer, and J. Gratch. The effectiveness of social influence tactics when used by a virtual agent. In *Proceedings of the ACM International Conference on Intelligent Virtual Agents*, pages 22–29, 2019.

[48] R. Mayer. *Multimedia learning*. New York, NY: Cambridge University Press, 2001.

[49] R. Mayer, K. Sabko, and P. Mautone. Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, 95:419–425, 2003.

[50] R. Moreno. *Multimedia learning with animated pedagogical agents*, pages 507–523. New York, NY: Cambridge University Press, 2005.

[51] R. Moreno, R. Mayer, H. Spires, and J. Lester. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 521:177–213, 2001.

[52] M. Morris and D. Keltner. How emotions work: An analysis of the social functions of emotional expression in negotiations. *Research in Organizational Behavior*, 22:1–50, 2015.

[53] G. Norman. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, 15(5):625–632, 2010.

[54] N. Norouzi, G. Bruder, B. Belna, S. Mutter, D. Turgut, and G. Welch. *Transactions on Computational Science and Computational Intelligence: Artificial Intelligence in IoT*, chapter A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things, pages 1–24. Springer, Cham, 2019.

[55] N. Norouzi, K. Kim, J. Hochreiter, M. Lee, S. Daher, G. Bruder, and G. Welch. A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In *Proceedings of the International Conference on Intelligent Virtual Agents*, pages 17–22, 2018.

[56] C. Opfermann, K. Pitsch, R. Yaghoubzadeh, and S. Kopp. The communicative activity of 'making suggestions' as an interactional process: Towards a dialog model for HAI. In *Proceedings of the International Conference on Human Agent Interaction*, pages 161–170, 2017.

[57] T. Piumsomboon, G. Lee, J. Hart, B. Ens, R. Lindeman, B. Thomas, and M. Billinghurst. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 46:1–13, 2018.

[58] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, New York, NY, USA, 1996.

[59] S. Robertson, R. Solomon, M. Riedl, T. W. Gillespie, T. Chociemski, V. Master, and A. Mohan. The visual design and implementation of an embodied conversational agent in a shared decision-making context (eCoach). In *International Conference on Learning and Collaboration Technologies*, pages 427–437, 2015.

[60] N. Schroeder and O. Adesope. A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education*, 46:229–251, 2014.

[61] A. Shamekhi, Q. Liao, D. Wang, R. Bellamy, and T. Erickson. Face Value? Exploring the effects of embodiment for a group facilitation agent. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[62] J. Short, E. Williams, and B. Christie. *The social psychology of telecommunications*. London: Wiley, 1976.

[63] C. Stephanidis, G. Salvendy, J. Chen, J. Dong, V. Duffy, X. Fang, C. Fidopiastis, G. Fragomeni, L. Fu, Y. Guo, D. Harris, A. Ioannou, K. Jeong, S. Konomi, H. Krömker, M. Kurosu, J. Lewis, A. Marcus, G. Meiselwitz, A. Moallem, H. Mori, F. Nah, S. Ntoa, P.-L. Rau, D. Schmorrow, K. Siau, N. Streitz, W. Wang, S. Yamamoto, P. Zaphiris, and J. Zhou. Seven HCI Grand Challenges. *International Journal of Human–Computer Interaction*, 35(14):1229–1269, 2019.

[64] G. Van Kleef, C. De Dreu, and A. Manstead. An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, 42:45–96, 2010.

[65] S. Van Mulken, E. Andre, and J. Muller. *The persona effect: How substantial is it?*, pages 53–66. Berlin, Germany: Springer, 1998.

[66] J. B. Walther. Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19:52–90, 1992.

[67] I. Wang, J. Smith, and J. Ruiz. Exploring virtual agents for augmented reality. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 281:1–12, 2019.

[68] G. Welch, G. Bruder, P. Squire, and R. Schubert. Anticipating Widespread Augmented Reality: Insights from the 2018 AR Visioning Workshop. Technical report, University of Central Florida and Office of Naval Research, August 2019.

[69] Z. Zhang and T. W. Bickmore. Medical shared decision making with a virtual agent. In *Proceedings of the International Conference on Intelligent Virtual Agents*, pages 113–118, 2018.