IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 26 June 2024, accepted 17 September 2024, date of publication 11 October 2024, date of current version 24 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3478751

SURVEY

Perceived Intelligence in Human–Robot Interaction: A Review

INARA TUSSEYEVA[1,2], ANARA SANDYGULOVA[3], AND MATTEO RUBAGOTTI[3], (Senior Member, IEEE)

[1]Department of Intelligent Systems and Cybersecurity, Astana IT University, Astana 010000, Kazakhstan
[2]Institute of Smart Systems and Artificial Intelligence, Nazarbayev University, Astana 010000, Kazakhstan
[3]Department of Robotics and Mechatronics, School of Engineering and Digital Sciences, Nazarbayev University, Astana 010000, Kazakhstan

Corresponding author: Anara Sandygulova (anara.sandygulova@nu.edu.kz)

The work of Anara Sandygulova was supported by under Project 11022021CRP1502. The work of Matteo Rubagotti was supported by Nazarbayev University under the Faculty Development Competitive Research Grants Program for 2024–2026 under Grant 201223FD8808.

ABSTRACT The aim of this paper is to provide a survey on the perception of robot intelligence. After a general overview, the paper specifically focuses on how perceived intelligence varies either between before and after the experiment ("pre-test–post-test variation"), and in subsequent sessions due to habituation. After reviewing the main characteristics of autonomous agents and robots that have been shown in the literature to influence the perception of robot intelligence, papers focusing on the variation in time of perceived intelligence are analyzed in detail. Even if no unanimous conclusion was reached in the literature, evidence suggests that, in general, when a significant variation is detected, perceived intelligence tends to increase from pre-test to post-test evaluations when commercial or more recent robot platforms are used, while it tends to decrease in the case of custom-made or less recent robots. On the other hand, when a significant variation is detected, perceived intelligence seems to increase due to habituation.

INDEX TERMS Habituation, human–robot interaction, perceived intelligence, pre-test–post-test variation.

I. INTRODUCTION

The field of human-robot interaction (HRI) has rapidly developed during past years, with the aim of providing robots with hardware, software, sensors and algorithms that can ensure comfortable, reliable and transparent communication with humans (see, e.g., [1], [2], [3]). The applications of HRI span several domains, such as manufacturing [4], service robotics [5], [6] and agriculture [7]. At the same time, many different aspects of HRI have been studied, such as safety [8], trust [9], perceived safety [10] and communication [11].

An important aspect of HRI is *perceived intelligence*, i.e., the subjective evaluation formed by human subjects regarding the cognitive abilities exhibited by the robot. High levels of perceived intelligence, which usually correspond to high levels of trust, are important to make robots accepted by potential users. Perceived intelligence has been studied not only for robots, but for a wider category of intelligent agents, such as chatbots and smartphones. In [12] it is mentioned that, if an agent is perceived as more intelligent, its perceived usefulness increases [13], together with the trust towards it [14].

If perceived intelligence influences perceived usefulness and trust, it is in turn influenced by different factors, as studied in [15], [16], [17], and [18]. According to the categorization in [12], the perceived intelligence of an agent increases with the following characteristics:

- *Autonomy:* the extent to which an agent can operate in an independent and goal-directed way without human intervention.
- *Adaptability:* the ability to improve the match between its functioning and the environment.
- *Reactivity:* the capacity to react to changes in the environment.
- *Multifunctionality:* the ability of a single agent to execute multiple functions.
- *Cooperativeness:* the capability to cooperate with other agents to achieve a common goal.

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Loconsole.

- *Human-like interaction:* the degree to which an agent communicates and interacts with humans in a natural, human-like fashion.
- *Personality:* the ability to show the properties of a credible character.

In addition to the above-mentioned general results, the factors that influence perceived intelligence were also studied specifically for robots. To the best of the authors' awareness of the literature, perceived intelligence is positively influenced by the following factors:

- *Task Performance*: Robots that effectively solve the assigned tasks tend to be seen as more intelligent. This concept is related to the concept of trust, i.e., how much a user can be confident that the robot will successfully complete an assigned task [19], [20].
- *Transparency*: The task assigned to the robot is not always clearly specified to the human, and this influences how a user assesses task performance. Therefore, providing comprehensive information regarding the robot behavior before or during the interaction [21] typically makes the robot more understandable and trustworthy for humans. It is worth mentioning that the term "transparency", in the field of artificial intelligence (AI), is related to the concepts of interpretability [22] and explainability [23] of AI processes, methods and functions. These capabilities of AI systems enable humans to grasp the decision-making processes behind their complex models [24].
- *Animacy*: This concept refers to the ability of the robot to be (quoting from [25]) "perceived as entity that exhibits lifelike traits or has a lifelike appearance" [25], [26], [27]. Indeed, "being alive is one of the major criteria that distinguish human beings from machines, but since robots exhibit movement and intentional behaviour, it is not obvious how human beings perceive them" [28]. The link with perceived intelligence is given by the fact that intelligence is instinctively attributed by humans to other living beings.
- *Appearance and Anthropomorphism*: Human beings are used to expect other human beings to be more intelligent than, for example, animals or machines. As also mentioned in [29], "people expect a humanoid robot to interact with a human-level of intelligence, whereas a robotic dog is expected to have the lower-level of intelligence of a dog". Therefore, a higher level of intelligence is typically assigned to robots that appear more human-like [25], [26], [30], [31], [32], [33]. Indeed, anthropomorphism (i.e., the tendency to attribute human-like qualities to non-human entities) and perceived intelligence are known to be positively correlated [34].
- *Human-Like Gestures*: If the robot makes gestures typically associated with human beings, then it is often evaluated as more intelligent [35], [36]. This result is related to that observed for anthropomorphism, although in this case the robot motion, rather than its appearance, is taken into account.
- *Social Interaction and Voice*: Finally, as mentioned in [37], robots are perceived as more intelligent if they can recognize, adapt to, and predict human behaviors. In case a robot can speak, voice tone and speed of talking also influence perceived intelligence [36], [38], [39], [40].

Thousands of papers were written in recent years in which the perceived intelligence of robots was assessed, but, to the best of the author's knowledge, no review papers were written to summarize these results. Therefore, after having surveyed the main characteristics related to perceived intelligence in the previous part of this introduction, the main contribution of this review relies on understanding how perceived intelligence changes, either (i) between before and after interacting with a robot, or (ii) as a result of multiple interactions with it. A better understanding of this aspect can be instrumental to design robotic systems that are still perceived as trustworthy and useful after interacting with them. To refer to case (i), the term *pre-test–post-test variation* will be employed, implying that the variation concerns perceived intelligence. The term "pre-test–post-test" has been used in different research fields (e.g., education [41] and medicine [42]) to refer to the measurement of a dependent variable before and after an experimental procedure. In this paper, the term "pre-test–post-test variation" refers to the difference between the expected robot intelligence (typically based on either robot appearance or task description, or both) and its perception after interacting with the robot. In case (ii),the term *habituation* is used. This concept has already been studied in HRI to understand the effects of human participants becoming accustomed to interacting with robots (see, e.g., [43], [44]). Regardless of its specific contextualization in HRI, habituation was first defined in the field of neurobiology as "a behavioral response decrement that results from repeated stimulation and that does not involve sensory adaptation/sensory fatigue or motor fatigue" [45].

The remainder of the paper is structured as follows: Section II describes general aspects of the surveyed papers, such as the methods (questionnaires) used for perceived intelligence assessment and a description of the robots used in the experiments. Section III reports a brief survey of existing papers, with the objective of understanding how either pre-test–post-test variation or the effect of habituation was studied in relation to perceived intelligence. Section IV provides a discussion of the results obtained in the surveyed papers, with the aim of detecting general trends in the variation of perceived intelligence. Finally, thoughts regarding future research are briefly reported in Section V.

II. OVERVIEW OF THE SURVEYED PAPERS

The publications studying the variation of perceived intelligence analyzed in this survey correspond to [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68],

[69], and [70], listed in chronological order, which will be analyzed in detail in the remainder of the paper. In [46], [47], [55], [59], [61], [62], [63], [65], [67], and [69] perceived intelligence was rated after each of multiple experimental sessions, possibly assessing the effect of habituation. In [48], [49], [50], [51], [52], [53], [56], [57], [58], [66], [68], and [70], instead, authors concentrated on pre-test–post-test variation. Articles [54], [60], [64] focused on both topics. In the remainder of this section, we provide information on (A) how perceived intelligence was assessed, (B) what robots were used for the experiments, and (C) how the chosen papers were selected.

A. ASSESSMENT METHODS FOR PERCEIVED INTELLIGENCE

Since its introduction in 2009 by Bartneck et al., the Godspeed questionnaire [28] has been the most commonly used assessment method to measure perceived intelligence – among other characteristics – of robots. The ideas at the basis of this questionnaire were derived from Warner and Sugarman's *intellectual evaluation scale* [71]. Specifically, the *perceived intelligence* scale consists of the evaluation of the following robot capabilities in the form of five-point semantic differential scales with Likert type scaling: *incompetent-competent*, *ignorant-knowledgeable*, *irresponsible-responsible*, *unintelligent-intelligent*, *foolish-sensible*. It is worth mentioning that Bartneck and co-authors had already mentioned the use of a robot intellectual evaluation scale, excluding the *incompetent-competent* item, prior to [28], and precisely in [25], [26], [72], [73], and [74]. The Godspeed questionnaire was used in [48], [49], [50], [51], [53], [54], [57], [58], [59], [61], [62], [63], [65], and [66]. It is also pertinent to note that this questionnaire was analyzed by several researchers and some inconsistencies were found [34], [75], [76]. More specifically, the rates of anthropomorphism, animacy, likeability, and perceived intelligence were closely correlated, showing inappropriateness in use as distinct concepts for evaluation [34]. However, the internal reliability of perceived intelligence rating was overall high.

A concept that can be considered as practically equivalent to perceived intelligence is that of *competence* of a robot, as perceived by human subjects. Indeed, *incompetent-competent* is one of the Godspeed questionnaire items, and the term *perceived intelligence* will be used in the remainder of the paper to refer to the concept of *competence* as well. Competence is used as a general category in the Robotic Social Attributes Scale (RoSAS) [76] and is further split into the evaluation, via a Likert scale, of how well the following adjectives (items) describe the robot: *reliable, competent, knowledgeable, interactive, responsive* and *capable*. Specifically, competence according to the RoSAS questionnaire was studied in [55], [56], [60], [64], and [70].

Ad-hoc questionnaires were also employed. A questionnaire that, similarly to RoSAS, evaluated competence, was used in [47], by assessing (again via a Likert scale) the perception of the following robot qualities: *intelligent, organized, expert* and *thorough*. The other questionnaire with the competence rate was used in [68] where 5 items *competent, confident, independent, competitive, intelligent*, provided by [77], were rated with a 7-point Likert scale. A partial version of the perceived intelligence evaluation in the Godspeed questionnaire was instead used in [52], including only the following items: *unintelligent-intelligent*, *incompetent-competent*. In [46], participants were asked the following questions: "How intelligently did the robot respond to your directions in each trial?" and "In which trial did the robot behave the most intelligently?". These two questions were rated on a Likert scale from 1 ("very intelligent") to 5 ("not at all intelligent"). Finally, in [69], a 5-point semantic differential scale questionnaire was used with the following question: "Please rate your impression of the robot on these scales:" from "non-intelligent" to "intelligent".

B. ROBOTS USED IN THE EXPERIMENTS

The most commonly used types of robots in the analyzed works were humanoid robots: Nao in [49], [51], [53], and [62], Pepper in [58], [63], and [66], the 3D blended embodiment Furhat in [56], [60], and [64], the JAMES robot bartender system (iCat with torso and arms) in [48], [51], and [52], the humanoid robot Rapiro in [65] and Ray (RoboThespian 4) in [68], a female version of the Android Actroid-F in [50], Emys in [57] and Robi in [54]. A robot prototype with the moving base and the screen on the top was utilized in [70]. The combination of humanoid and mobile robots was studied in [59] with Nao and Roomba, in [46] with four appearance configurations of PeopleBot, and in [67] with QTrobot and Mistry robot. The collaborative manipulators KUKA LBR iiwa 7 R800, Universal Robots UR5 and Kinova Gen3 were used in [55], [61], and [69], respectively. Finally, the two virtual agents IVA Vince (robot-like) and IVA Billie (human-like) were compared in [47].

C. RESEARCH METHOD

In order to select the papers for this survey, we first used different search engines (Google Scholar, IEEEXplore, ScienceDirect, Scopus) for the keywords "robot" AND ("perceived intelligence" OR ("competence" AND "RoSAS")) AND ("pre- and post-" OR "pre-test–post-test" OR "habituation" OR "longitudinal" OR "novelty effect" OR "ordering effect"). Indeed, the term *competence* is often used with different meanings outside the RoSAS questionnaire, and terms such as *longitudinal study, novelty effect* and *ordering effect* were often used to indicate a study of habituation, even when the term "habituation" was not explicitly mentioned. In addition to the publications found with this method, we also considered works cited in them that were potentially relevant. Overall, this led to approximately 400 papers. These were further screened by eliminating those entries which did not satisfy the following requirements:

- The work is published in English language either as an international journal papers, in the proceedings of an

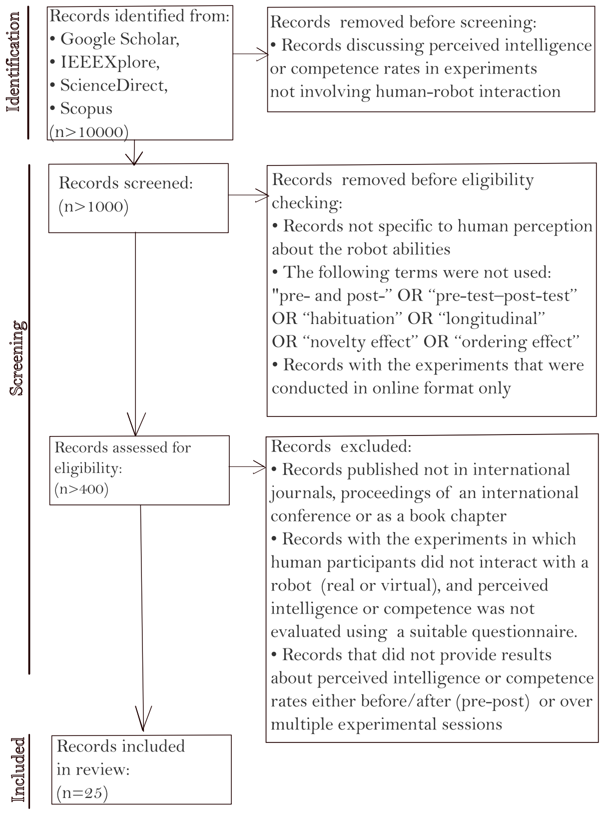

FIGURE 1. Flow diagram of the literature search and screening process.

international conference or as a book chapter.

- Experiments are run in which human participants interact with a robot (either real or, in few cases, virtual), and perceived intelligence is evaluated using a suitable questionnaire.
- Either pre-test–post-test variation or the change in perceived intelligence over multiple sessions is assessed and presented in the results.

After this screening, the above-mentioned 25 works [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70] were obtained. The flow chart of the searching and screening process is shown in Fig. 1.

III. ANALYSIS OF THE SURVEYED PAPERS

In this section we provide a description of the surveyed papers in chronological order, by focusing either on pre-test–post-test variation or on the connection between habituation and perceived intelligence, both in the description of the experiments and in the interpretation of the questionnaire results by the authors. Additional experimental tasks and/or questionnaires not related to perceived intelligence are mostly not reported, to focus on the survey topic. It is worth noticing that in [46], [53], [58], [60], [61], [63], [64], [65], and [68] the so-called *Wizard of Oz* approach [78] was used, meaning that these robots were partially controlled by a human operator without participants being aware of it.

Perceived robot intelligence was rated in [46] by having participants play the so-called "hot and cold game" with a robot. In this game, the robot searched for an object randomly chosen by the participant in a room, while the participant guided the robot by saying "hot" or "cold", respectively, if it was either close or far from the object. The experiment consisted of three rounds, in which the robot would always avoid obstacles in the room: during the first and third rounds the robot reacted to human words by changing its motion direction on the word "cold"; in the second round, instead, it moved randomly across the room without reacting to human words. Brief contact sessions and a relatively small number of participants were identified by the authors as the main limitations of their work. The results of the questionnaire showed that the robot was rated as more intelligent during the third round of the trial, whereas the lowest marks were given in the second round. This work did not assess habituation by studying the variation of perceived intelligence while the same robot behavior was observed; therefore, it is not possible to conclude if habituation had a positive or negative effect on the perception of intelligence. However, the different evaluation of the same robot behavior in rounds one and three seems to indicate that perceived intelligence is influenced by what participants experienced immediately before the round. More specifically, an improvement of the actual robot intelligence led to a positive bias in its perception.

In [47], each participant interacted with one of two virtual agents: a human-like agent with child voice named Billie and a robot-like one with machine voice named Vince; each of them, throughout the whole experiment for a single participant, would either make gestures when speaking (co-speech gestures), or not (the type of agent and the presence/absence of gestures were randomly assigned). The virtual agent performed a self-introduction and asked the participant to answer a questionnaire about perceived level of competence. After filling the questionnaire, the participant watched a presentation given by the virtual agent, before evaluating competence once more. For both agents it was observed that, on average, competence rates improved in the second questionnaire when the use of co-speech gestures was present, and worsened when it was absent. Although the robot behavior between the first session (introduction) and the second session (presentation) was different, to the best of our understanding this behavior did not display a varying level of intelligence. Therefore, the variation of perceived intelligence can be attributed to habituation, differently from [46], in which the robot behavior was purposefully changed to be less intelligent in the second session. The authors of [47] stated that a limitation of their study is the fact that it did not consider the influence of other elements, such as the agent's voice or other nonverbal cues.

The research described in [48] assessed how participants rated the intelligence of a bartender robot that served them drinks by following different behaviors, with different levels of social interaction, and with different duration of the experiment. The rating was obtained both before and after the experiment. Participants filled the pre-experiment questionnaire after being told they would interact with a bartender robot, but this had not been shown to them yet. Regardless of the employed behavior, it was concluded that: (i) the users' expectations about robot intelligence were higher than the perceived intelligence rate obtained after interaction (although the significance of the related test was not studied), and (ii) the duration of the session had a positive impact on perceived intelligence, as evaluated after the experiment.

In [49], the Nao robot was standing in front of the participant and speaking several sentences that were, in different sessions for the same human subject, possibly accompanied by gestures and/or eye contact. Participants were tasked to remember the messages communicated by the robot and to repeat them at the end of the session. Perceived intelligence was evaluated both before the experiment (after providing a rather detailed description of the upcoming activity) and after each of the four sessions that composed it. In conclusion, no significant variation of perceived intelligence was detected between pre- and post-experiment questionnaires.

Human perception and trust towards a lifelike android robot were studied in [50], in which the robot engaged the participants in a small conversation, and asked them to carry out simple tasks, such as moving a small box on a table and touching its hand. The perceived intelligence rating was significantly lower after interacting with the robot compared to the pre-experimental questionnaire (before filling it, participants were shown pictures of the robot).

A multi-party social interaction was the focus of [51], in which a Nao bartender robot interacted with two users by talking to them and serving them drinks. The experiment consisted of four sessions in which learning-based and hand-coded algorithms were used two times each. Perceived intelligence was evaluated both before the experiment and at the end of all four sessions. The users rated the robot as significantly more intelligent before the experiment compared to the assessment obtained after four interactions. This result was explained by the authors with the limited task domain supported by the robot. Indeed, "although in real bar situations, it seems perfectly reasonable to assume that a customer can order without the bartender explicitly asking, in this more artificial human-robot interaction setting, this strategy might have been too confusing, resulting in the lower scores presented above" [51].

The authors of [52] extended their research [48] by proposing an uncertainty-aware algorithm to determine the actions of their bartender robot in a similar context as that of [48]. In spite of the more complex algorithm, pre-test–post-test results confirmed the results of [48], i.e., perceived intelligence decreased post-test (although this difference was not significant); this was true for all participants with and without prior experience with the same experimental setup. The authors concluded that "people's expectations of a robot's interactive capabilities tend to outstrip their actual experience of interacting with it, even when they have previous experience with the same robot" [52].

The authors of the study conducted in [53] investigated how varying levels of social verification in robot behavior influenced human perceptions of the robot. They examined two primary behaviors: "idle," representing a low level of social verification, and "meaningful," representing a high level. A motionless robot was always presented first as a baseline condition. Subjects were randomly assigned to one of the social verification conditions. In these conditions, the robot directed participants on which item to take from a box and where to place it, using one of the specified behaviors. Each condition included three different robot motions, with one experimental set comprising four tasks in total. There were four experimental sets overall. After completing each set, participants filled out a questionnaire. Results showed that the robot was perceived as more intelligent when performing in the "meaningful" condition compared to the "idle" condition. Furthermore, the study found a significant increase in perceived intelligence from the baseline to the subsequent three motions in the "meaningful" condition, whereas the "idle" condition showed no significant differences.

In [54], human subjects engaged in pairs – either actively or passively – in conversations with a small humanoid robot over two sessions. Perceived intelligence was assessed before starting the experiment, after the first session (in which its average rating underwent a marginally significant decrease compared to the pre-experiment questionnaire) and after the second session (in which its rating increased, though not significantly, with respect to the first session but without reaching the pre-experiment level). According to the authors, the lowering in perceived intelligence after the first session "could be attributed to the experimental setting", as "people were instructed to interact with the robot in a fixed protocol" that limited the complexity of the interaction compared to their initial expectations. Also, the authors observed that for the android robot in [50] there was a sharper drop of perceived intelligence as compared to the humanoid robot of [54]; the provided explanation was that the experiments in [50] were probably influenced by the so-called *uncanny valley* effect [79], which made humans dislike the android (and, in this case, also rating it as less intelligent) after interacting with it. Among the limitations of [54], the authors listed the cultural background of the participants, laboratory conditions and social desirability bias.

When evaluating human-to-robot handovers, the authors of [55] varied three factors – initial position, retraction speed, and grasp type – for a total of 8 conditions per user. The expectations about the results of competence levels among grasp types were not met by the researchers. The results showed significantly higher rates of competence for one type of grasp (namely, "quick grasp") compared to another type ("mating grasp"). Regardless of the obtained results relating competence to three factors, the authors examined whether the participants' perception changed over the course of repeated handover interactions with the robot, concluding that there was no significant trend in competence variation.

In [56], participants were gradually exposed to the motionless 3D blended embodiment Furhat (consisting of a robotic head) with different variations in terms of human-likeness. There were three stages of interaction: in what can be considered as a pre-session experience, the participants only observed the robot when this was motionless, just blinking the eyes; then, the robot introduced itself by speaking without reacting to the participant's actions; finally, the robot interacted with the participant by playing a game involving a conversation. While there was no significant difference in robot competence ratings between the pre-session experience and the case when the robot was only speaking, there was a significant difference in perceived competence between the last stage compared to the case when the robot was only speaking, with this difference being more pronounced when the more human-like variants of the face were used. This confirms the foreseeable expectation that "the more participants are exposed to the robot's capabilities, the more competent they perceive it" [56]. Also, the pre-test-post-test variation was found to be either not significant or significantly increasing depending on the complexity of the robot behavior. Indeed, a more complex type of interaction led to a significant increase of competence compared to its pre-session evaluation.

In [57], the authors explored the impact of varying levels of robot assertiveness on decision-making processes and human perceptions within persuasive HRI. Participants, assuming the role of a country leader during wartime, crafted a narrative while receiving counsel from two identical-looking robots, Emys and Glin. The experimental setup involved two conditions: both robots behaving neutrally, or exhibiting contrasting assertiveness, with Emys showing high assertiveness and Glin low assertiveness. Participants assessed their perceptions of the robots through questionnaires administered both before and after the experiment. The findings indicated no significant changes in perceived intelligence or competence when the robots displayed different levels of assertiveness. However, a notable difference in perceived competence of Glin was observed when both robots behaved neutrally. The study did not provide specific outcome values for the neutral condition.

In [58], teams of two participants completed a set of five one-minute rounds collaborating on a story-making task with a Pepper robot that was either positive (i.e., joyful) or neutral in its voice and spoken utterances. An increase of perceived intelligence ratings between pre- and post-questionnaire was observed in both cases, but the increase was significant only in the neutral robot condition. Several limitations were listed by the authors, including the absence of previous experience of the majority of the participants interacting with humanoid robots.

With the aim to evaluate perceived intelligence of a vacuum cleaner Roomba robot during its interaction with a humanoid Nao robot, the authors of [59] conducted a within-subject design study with two conditions: "interactive" and "no interactive". The difference between these conditions was in the behavior of the Nao robot: it was either interacting with the Roomba robot or not. In the "interacting" condition, the Roomba robot consistently achieved a higher rating of perceived intelligence by the participants as compared to the "no-interactive" condition. However, there was no ordering effect for perceived intelligence.

In [60] and [64] participants played a geography-themed cooperative game with a Furhat robot three times (three to ten days apart) with a different robot embodiment (human-like, machine-like and morphed, i.e., a mix of human-like and machine-like) and had a social face-to-face chat before and after the game. Data from study [60] showed that the human-like robot face was perceived as significantly more competent compared to the morphed face, while the findings from study [64] additionally suggest significant differences in competence ratings between human-like and machine-like embodiments. In both works, no significant changes in perceived intelligence between pre- and post-questionnaires were found.

In [61] it was investigated how animal-like character animation principles (appearance, smooth motion, breathing, gaze and posture) can enhance perceived intelligence of the Universal Robots UR5 cobot in a collaborative task. Results from two user studies suggested that, while appearance, smooth motion and posture did not have an effect on cobot's perceived intelligence, the presence of animal-like breathing motions as well as gazing behaviour significantly improved perceived intelligence. However, no correlation was found between these factors and any ordering effects.

In [62], a Nao robot interacted with participants by means of spoken sentences and gestures. The interaction during each of two sessions consisted of ten sentences pronounced by the robot with their accompanying gestures. For half of the participants, in the first session the first five sentences were determined using a pre-programmed method, while the next five sentences were selected using an ad-hoc machine-learning-based method developed in the paper; the reverse order was followed in the second session. For the other half of the participants, the opposite sequence of methods was followed through the sessions. While no significant differences on perceived intelligence were found regarding the use of different methods, it was determined that, regardless of the used order of methods, perceived intelligence significantly increased from the first session to the second.

The goal of the research reported in [63] was to evaluate a 5-week mindfulness sessions administered by either a teleoperated Pepper robot controlled by an experienced human coach via teleoperation, or by the same human coach directly. In both cases, no significant differences were observed in perceived intelligence over time. This was expected by the authors as the robot behavior and appearance did not change.

The focus of the research work [65] was to study the perception of the fact that a Rapiro robot would take (or not) the initiative to initiate an interaction with human participants. Participants interacted in sequence with both robot conditions: active (the robot was first to greet participants) and passive (the robot waited for participants to greet it first), with the order being counterbalanced. Neither the specific condition (active or passive) nor the order of execution of the conditions led to any significant result in terms of variation of perceived intelligence. The limitations of the conducted research were related to the small sample size and variety of participants' cultural backgrounds.

In [66] the study investigated how the implementation of an overt self-talk system in robots, designed to mimic human-like inner speech, would influence participants' trust and perception. The research included both pre-test and post-test experiments. Participants were tasked with placing utensils appropriately on a virtual table according to etiquette rules, with the robot assisting them. In the experimental group, the robot not only interacted with the participants but also engaged in self-talk, while participants in the control group used only outer speech system. Results indicated that participants in the experimental group rated the robot's perceived intelligence higher after the interaction compared to those in the control group. The authors anticipated this overall increase in positive perception following the interaction with the robot.

The work described in [67] explored how the employees of a tech company perceived a QTrobot and a Misty robot as well-being coaches programmed to deliver positive psychology exercises four times over four weeks. No significant differences in competence ratings were obtained neither between the two robots, nor between subsequent sessions.

The main objective of the article [68] on robot competence was to answer the research question: whether watching positive or negative depictions of robots in the media can influence human perceptions of a real-life social robot's abilities. To explore this, the researchers conducted a two-phase experiment. In the first phase, participants completed an online pre-session survey to assess their general perceptions of robots. In the second phase, the same participants took part in an in-person lab session. During this session, participants first watched media stimuli and recognized films and robots from short clips. Then, they observed a conversation between a robot and an experimenter and answered questions about the robot's capabilities. Participants were divided into two groups. One group watched videos where robots were depicted as trustworthy, empathetic, warm, and unthreatening (positive video condition). The other group watched videos where robots were portrayed as cold, dangerous, threatening, and untrustworthy (negative video condition). The results showed no significant differences in perceived competence between the pre-session and post-session ratings across both groups. The authors suggested that future research might include behavioral assessments, implicit measurements, and longer interaction periods to gain deeper insights.

In [69], a collaborative manipulator executed a pick-and-place task while sharing its workspace with the human participant, who executed an independent task. Four different motion planning algorithms (ensuring human safety according to industrial standards) were used in four subsequent sessions of four minutes each, with their order being counterbalanced. The authors noticed that perceived intelligence increased significantly from the first session to the second; it then continued to increase during the next two sessions, but not significantly.

The study in [70] investigated the effect of robot anthropomorphic features on human perception of service robots. Two types of robot appearances were tested: human-like (with a head-screen displaying eyes, a smile, and a nose) and machine-like (with only a dot on the head-screen). Each participant interacted with one type of robot. The experiment followed a specific procedure: participants observed a motionless robot and then answered a questionnaire (visual contact only phase), then participants stood in front of the robot, listened to its speech about its characteristics and limitations, and then answered another questionnaire (visual and auditory contact phase), and finally participants moved along a predetermined course with the robot twice, answering a questionnaire (navigation phase). The results showed that the robot was rated significantly higher for competence after the navigation task compared to both the visual contact and the combination of visual and auditory contacts phases. The initial observation phase was considered a pre-interaction session, leading to the conclusion that robot competence perception increased after interaction with the robot. The authors suggested future research directions, including collecting additional data from participants and studying how potential errors made by the robot and the robot's gender characteristics might affect human perception.

IV. DISCUSSION

A summary of the key points of the surveyed papers is reported in Table 1. In particular, the table contains information on authors, year of publication, whether the paper analyzes the change of perceived intelligence in the case of pre-test-post-test variation or habituation, the number of experimental sessions, the number of participants, the type of employed questionnaire, the possible use of the Wizard of Oz approach, the obtained p-values related to the statistical analysis of the experimental data, and the corresponding conclusions of the significance of the obtained results.

The discussion of these results is provided separately for pre-test-post-test variation and habituation in the following. Notice that some papers are mentioned in more than one category: this is due to the same paper reporting different results for different sub-parts of the overall experimental activity described in it.

A. PRE-TEST–POST-TEST VARIATION

The results on pre-test-post-test variation are further summarized in Fig. 2. Paper numbers are reported on the horizontal

TABLE 1. Main characteristics of the surveyed papers, sorted by year of publication.

Article	Authors	Year	Case	Sessions number	Subjects number	Questionnaire	Wizard of Oz	p value	Significance of results
[47]	Bergmann, Eyssel, Kopp	2012	Habituation	2	80	Ad-hoc	No	$p = .04$	Either significant increase or significant decrease
[48]	Giuliani, Petrick, Foster, Gaschler, Isard, Pateraki, Sigalas	2013	Pre-test-post-test	1	40	Godspeed	No	No reported value	No reported quantitative results
[49]	Van Dijk, Torta, Cuijpers	2013	Pre-test-post-test	4	19	Godspeed	No	No reported value	No significant change
[50]	Haring, Matsumoto, Watanabe	2014	Pre-test-post-test	3	55	Godspeed	No	$p < .001$	Significant decrease
[51]	Keizer, Kastoris, Foster, Deshmukh, Lemon	2014	Pre-test-post-test	4	48	Godspeed	No	$p < .05$	Significant decrease
[52]	Keizer, Foster, Gaschler, Giuliani, Isard, Lemon	2014	Pre-test-post-test	4	24	Ad-hoc	No	No reported value	No significant change
[53]	Cuijpers, Knops	2015	Pre-test-post-test	4	73	Godspeed	Yes	meaningful case: $p = .02$, idle case: $p = .83$	Either significant increase or no significant change
[54]	Haring, Watanabe, Silvera-Tawil, Velonaki, Takahashi,	2015	Pre-test-post-test and habituation	3	42	Godspeed	No	$p = .06$ for pre-test-post-test variation, no reported value for habituation	No significant change
[55]	Pan, Croft, Niemeyer	2018	Habituation	8	22	ROSAS	No	No reported value	No significant change
[56]	Paetzel, Castellano	2019	Pre-test-post-test	3	48	ROSAS	No	$p = .002$	Significant increase
[57]	Paradeda, Ferreira, Oliveira, Martinho	2019	Pre-test-post-test	1	61	Godspeed	No	No reported value	No significant change
[58]	Rhim, Cheung, Pham, Bae, Zhang, Townsend, Lim	2019	Pre-test-post-test	5	78	Godspeed	Yes	neutral voice case: $p = .018$, no reported value in other cases	Either significant increase or no significant change
[59]	Ueno, Hayashi, Mizuuchi	2019	Habituation	2	22	Godspeed	No	no reported value	No significant change
[60]	Paetzel, Perugia, Castellano	2020	Pre-test-post-test and habituation	3	49	ROSAS	Yes	$p = .179$ for habituation, no reported value for pre-test-post-test variation	No significant change for both pre-test-post-test and habituation
[61]	Terzioglu, Multu, Sahin	2020	Habituation	2	72	Godspeed	Yes	no reported value	No significant change
[62]	Xiao, Silvera-Tawil, Pagnucco	2020	Habituation	2	12	Godspeed	No	$p = .023$	Significant increase
[63]	Bodala, Churamani, Gunes	2021	Habituation	5	18	Godspeed	Yes	no reported value	No significant change
[64]	Perugia, Paetzel-Prusmann, Alanenpaa, Castellano,	2021	Pre-test-post-test and habituation	3	52	ROSAS	Yes	$p = .057$	No significant change for both pre-test-post-test and habituation
[65]	Kan John, Zhu, Gedeon, Zhu	2022	Habituation	2	16	Godspeed	Yes	no reported value	No significant change
[66]	Pipitone, Geraci, D'Amico, Seidita, Chella,	2023	Pre-test-post-test	1	51	Godspeed	No	$p < .05$ for over inner speech system, no significance for outer speech system	Either significant increase or no significant change
[67]	Spitale, Axelsson, Gunes	2023	Habituation	4	26	Godspeed	No	no reported value	No significant change
[68]	Stein, Banks	2023	Pre-test-post-test	2	77	Ad-hoc	Yes	no reported value	No significant change
[69]	Tusseyeva, Oleinikov, Sandygulova, Rubagotti	2023	Habituation	4	48	Ad-hoc	No	$p = .042$	Significant increase
[70]	Dautzenberg, Ladwig, Rosenthal, von der Putten	2024	Pre-test-post-test	3	64	ROSAS	No	$p = .007$	Significant increase

axis, whereas the vertical axis indicates if there were any (marginally) significant perceived intelligence variations detected, or not. The size of each circle is proportional to the number of participants, which is indicated inside the circle.

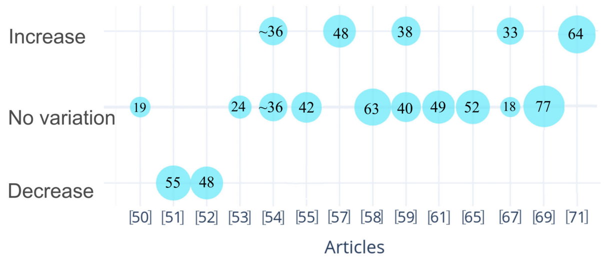

FIGURE 2. **Summary of the variation of perceived intelligence, in different papers in pre-test–post-test questionnaires.**

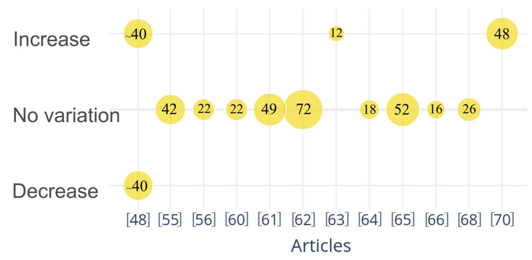

FIGURE 3. **Summary of the variation of perceived intelligence, in different papers due to habituation.**

There are several papers where two results were obtained for different experimental conditions. For instance, two different results for different sets of participants were obtained in [58]; as a consequence, the number of subjects for each of the two cases (38 with neutral robot and 40 with positive robot) was explicitly indicated in the paper.

No correlation was found between the obtained results and the number of participants. Between the pre- and the post-experiment questionnaires, the following results were reported on perceived intelligence:

- a significant increase in [53], [56], [58], [66], and [70];
- no significant variation in [49], [52], [53], [54], [56], [57], [58], [60], [64], [66], and [68];
- a significant decrease in [50] and [51].

A decreasing trend was also reported in [48], but, as the significance of the related test was not studied, this article is not reported in Fig. 2. In general, it was not possible to relate any specific characteristics of the robots or of the tasks to the obtained results. However, it appears that an increase of robot intelligence rate was observed when more commercially available research platforms were employed (the only exception was [70] , in which a robot prototype with moving base and screen was used). This kind of social robots are reliable and could be applied in real-world scenarios (see, e.g. [75], [80], [81], [82]). The interaction with them is largely predefined based on their appearance and capabilities. On the other hand, a decrease of perceived intelligence was observed in the experiments employing custom-built robots of the specific laboratory [48], [50], [51] .

Furthermore, as the papers in which a significant decrease of perceived intelligence was detected were published about ten years ago (in 2012 [47] and in 2014 [50], [51]), whereas the articles showing a significant increase of perceived intelligence were published more recently. A possible explanation is that earlier research might have had limited technological capabilities, whereas more recent research can take advantage of advanced capabilities such as autonomous robot behavior, speech recognition and computer vision.

B. EFFECT OF HABITUATION

A summary of the results on habituation analogous to that of Fig. 2 is reported in Fig. 3. As in the pre-test-post-test case, there are several papers where two results were obtained for different experimental conditions. For instance, out of the 80 participants of [47], some were randomly assigned to an agent making gestures (for which an increase was detected) and some to an agent not making gestures (for which a decrease was detected); this is the reason for the presence of two circles, for each of which we assume a size of approximately 40.

Habituation seemed to cause, regarding perceived intelligence,

- a significant increase in [47], [62], and [69];
- no significant change in [54], [55], [59], [60], [61], [63], [64], [65], and [67];
- a significant decrease in [47].

It is worth noticing that paper [46] could not be used to assess the effect of habituation as, despite running multiple experimental sessions each followed by a questionnaire, the robot behavior changed quite dramatically in subsequent sessions, and the order in which these behaviors were presented to the users was always the same. In particular, [46] presented a result on how the variation of perceived intelligence in subsequent sessions can be influenced by the previously-observed robot behavior. Specifically, after observing a non-intelligent robot behavior, participants tend to evaluate the robot as more intelligent when it behaves normally, compared to the same evaluation of this normal behavior without any previous experience; this observation should be accounted for in future research.

The great majority of surveyed papers showed no significant changes in perceived intelligence due to habituation. An increase was observed for a robot or virtual agent speaking with participants while making gestures in [47] and [62], and for a collaborative manipulator in [69]. However, again in [47] the perceived intelligence of the same virtual agent decreased when the latter was not making gestures. This can be related to the fact that, as already mentioned in the introduction, the presence of human-like gestures in general improves perceived intelligence ratings. Overall, we can conclude that habituation seems to have, on average, a positive effect on perceived intelligence. This general result is also supported by the conclusions of [48] in which, despite having a single experimental session, it was shown that a longer duration of the session corresponded to a higher rating of perceived intelligence.

A possible explanation of the general trend is that, as the robot behavior is complex, it is not easy to immediately infer the rules that determine its actions. Thus, as a better understanding of the robot behavior is achieved in subsequent sessions, the perceived intelligence rates also increase.

V. CONCLUSION AND FUTURE WORK

The obtained results seem to point to the decrease of perceived intelligence between pre- and post-experiment questionnaires, and to its increase in subsequent sessions due to habituation. However, these results are in no way conclusive. Indeed, no study was conducted so far focusing exclusively on these aspects, which have to be further investigated. A possibility would be to conduct experiments with different robots, different a-priori information given to participants and different tasks, to assess how perceived intelligence varies depending on transparency, animacy, human-like appearance and gestures.

In particular, the role of information (i.e., transparency) surely deserves further investigation. As for pre-test–post-test variation, one can expect that little information provided before the experiment on either robot appearance or task would more likely lead to a change in perceived intelligence as compared to the case in which a very detailed description is provided, or (even more) to the case when a video of the task is shown to the participant before the experiment. Regarding habituation, we can expect that either little initial information or a complex robot behavior that has to be understood would lead to a more significant variation of perceived intelligence, compared to the case in which participants know all details of the task beforehand, or in which the robot behavior is very easy to predict.

REFERENCES

[1] A. Zacharaki, I. Kostavelis, A. Gasteratos, and I. Dokas, "Safety bounds in human robot interaction: A survey," *Saf. Sci.*, vol. 127, Jul. 2020, Art. no. 104667.

[2] T. Law, M. Chita-Tegmark, and M. Scheutz, "The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study," *Int. J. Social Robot.*, vol. 13, no. 2, pp. 297–309, Apr. 2021.

[3] R. Jahanmahin, S. Masoud, J. Rickli, and A. Djuric, "Human–robot interactions in manufacturing: A survey of human behavior modeling," *Robot. Comput.-Integr. Manuf.*, vol. 78, Dec. 2022, Art. no. 102404.

[4] J. A. Marvel, S. Bagchi, M. Zimmerman, and B. Antonishek, "Towards effective interface designs for collaborative HRI in manufacturing: Metrics and measures," *ACM Trans. Hum.-Robot Interact.*, vol. 9, no. 4, pp. 1–55, Dec. 2020.

[5] H. Kim, K. K. F. So, and J. Wirtz, "Service robots: Applying social exchange theory to better understand human–robot interactions," *Tourism Manage.*, vol. 92, Oct. 2022, Art. no. 104537.

[6] C. S. Song and Y.-K. Kim, "The role of the human–robot interaction in consumers' acceptance of humanoid retail service robots," *J. Bus. Res.*, vol. 146, pp. 489–503, Jul. 2022.

[7] J. P. Vasconez, G. A. Kantor, and F. A. A. Cheein, "Human–robot interaction in agriculture: A survey and current challenges," *Biosystems Eng.*, vol. 179, pp. 35–48, Mar. 2019.

[8] S. Haddadin and E. Croft, "Physical human–robot interaction," in *Springer Handbook of Robotics*, 2016, pp. 1835–1874.

[9] C. S. Nam and J. B. Lyons, *Trust in Human-Robot Interaction*. New York, NY, USA: Academic, 2020.

[10] M. Rubagotti, I. Tusseyeva, S. Baltabayeva, D. Summers, and A. Sandygulova, "Perceived safety in physical human–robot interaction—A survey," *Robot. Auto. Syst.*, vol. 151, pp. 1–22, May 2022.

[11] A. Bonarini, "Communication in human–robot interaction," *Current Robot. Rep.*, vol. 1, pp. 279–285, Sep. 2020.

[12] M. Makkonen, M. Salo, and H. Pirkkalainen, "What makes a (ro)bot smart? Examining the antecedents of perceived intelligence in the context of using physical robots, software robots, and chatbots at work," in *Proc. Medit. Conf. Inf. Syst.*, 2022.

[13] S. Moussawi, M. Koufaris, and R. Benbunan-Fich, "How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents," *Electron. Markets*, vol. 31, no. 2, pp. 343–364, Jun. 2021.

[14] W. Kim, N. Kim, J. B. Lyons, and C. S. Nam, "Factors affecting trust in high-vulnerability human–robot interaction contexts: A structural equation modelling approach," *Appl. Ergonom.*, vol. 85, May 2020, Art. no. 103056.

[15] S. A. Rijsdijk and E. J. Hultink, "'Honey, have you seen our hamster?' Consumer evaluations of autonomous domestic products," *J. Product Innov. Manage.*, vol. 20, no. 3, pp. 204–216, 2003.

[16] S. A. Rijsdijk, E. J. Hultink, and A. Diamantopoulos, "Product intelligence: Its conceptualization, measurement and impact on consumer satisfaction," *J. Acad. Marketing Sci.*, vol. 35, no. 3, pp. 340–356, Sep. 2007.

[17] S. A. Rijsdijk and E. J. Hultink, "How today's consumers perceive tomorrow's smart products," *J. Product Innov. Manage.*, vol. 26, no. 1, pp. 24–42, Jan. 2009.

[18] W.-J. Lee and S. Shin, "Effects of product smartness on satisfaction: Focused on the perceived characteristics of smartphones," *J. Theor. Appl. Electron. Commerce Res.*, vol. 13, no. 2, pp. 1–14, May 2018.

[19] K. E. Schaefer, T. L. Sanders, R. E. Yordon, D. R. Billings, and P. A. Hancock, "Classification of robot form: Factors predicting perceived trustworthiness," in *Proc. SAGE Publications Hum. Factors Ergonom. Soc. Annu. Meeting*, 2012, vol. 56, no. 1, pp. 1548–1552.

[20] L. Graf, M. Torkar, E. Stückelmaier, R. Sichler, P. Malafosse, K. Fischer, and O. Palinko, "Perceived trustworthiness of an interactive robotic system," in *Proc. 17th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2022, pp. 773–777.

[21] L. Kluy and E. Roesler, "Working with industrial cobots: The influence of reliability and transparency on perception and trust," in *Proc. SAGE Publications Hum. Factors Ergonom. Soc. Annu. Meeting*, 2021, vol. 65, no. 1, pp. 77–81.

[22] B. Mittelstadt, "Interpretability and transparency in artificial intelligence," in *The Oxford Handbook of Digital Ethics*, vol. 20, 1093.

[23] N. Balasubramaniam, M. Kauppinen, K. Hiekkanen, and S. Kujala, "Transparency and explainability of AI systems: Ethical guidelines in practice," in *Proc. Int. Work. Conf. Requirements Eng., Found. for Softw. Qual.* Cham, Switzerland: Springer, 2022, pp. 3–18.

[24] B. Vyas, "Explainable AI: Assessing methods to make AI systems more transparent and interpretable," *Int. J. New Media Stud., Int. Peer Reviewed Scholarly Indexed J.*, vol. 10, no. 1, pp. 236–242, 2023.

[25] C. Bartneck, T. Kanda, O. Mubin, and A. A. Mahmud, "The perception of animacy and intelligence based on a robot's embodiment," in *Proc. 7th IEEE-RAS Int. Conf. Humanoid Robots*, Nov. 2007, pp. 300–305.

[26] C. Bartneck, T. Kanda, O. Mubin, and A. Al Mahmud, "Does the design of a robot influence its animacy and perceived intelligence?" *Int. J. Social Robot.*, vol. 1, no. 2, pp. 195–204, Apr. 2009.

[27] G. Schillaci, S. Bodiroža, and V. V. Hafner, "Evaluating the effect of saliency detection and attention manipulation in human–robot interaction," *Int. J. Social Robot.*, vol. 5, no. 1, pp. 139–152, Jan. 2013.

[28] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *Int. J. Social Robot.*, vol. 1, no. 1, pp. 71–81, Jan. 2009.

[29] S. Krening and K. M. Feigh, "Characteristics that influence perceived intelligence in AI design," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2018, pp. 1637–1641.

[30] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler, "All robots are not created equal: The design and perception of humanoid robot heads," in *Proc. 4th Conf. Designing Interact. Syst., Processes, Practices, Methods, Techn.*, Jun. 2002.

[31] B. R. Duffy, "Anthropomorphism and the social robot," *Robot. Auto. Syst.*, vol. 42, nos. 3–4, pp. 177–190, Mar. 2003.

[32] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robot. Auto. Syst.*, vol. 42, nos. 3–4, pp. 143–166, Mar. 2003.

[33] G. Trovato, J. G. Ramos, H. Azevedo, A. Moroni, S. Magossi, R. Simmons, H. Ishii, and A. Takanishi, "A receptionist robot for Brazilian people: Study on interaction involving illiterates," *J. Paladyn Behav. Robot.*, vol. 8, no. 1, pp. 1–17, Apr. 2017.

[34] C.-C. Ho and K. F. MacDorman, "Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices," *Comput. Hum. Behav.*, vol. 26, no. 6, pp. 1508–1518, Nov. 2010.

[35] H. Kose-Bagci, E. Ferrari, K. Dautenhahn, D. S. Syrdal, and C. L. Nehaniv, "Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot," *Adv. Robot.*, vol. 23, no. 14, pp. 1951–1996, Jan. 2009.

[36] A. Mileounis, R. H. Cuijpers, and E. I. Barakova, "Creating robots with personality: The effect of personality on social intelligence," in *Proc. Artif. Comput. Biol. Med., Int. Work-Conf. Interplay Natural Artif. Comput.* Cham, Switzerland: Springer, 2015, pp. 119–132.

[37] K. A. Barchard, L. Lapping-Carr, R. S. Westfall, A. Fink-Armold, S. B. Banisetty, and D. Feil-Seifer, "Measuring the perceived social intelligence of robots," *ACM Trans. Hum.-Robot Interact.*, vol. 9, no. 4, pp. 1–29, Dec. 2020.

[38] S. S. Sundar, E. H. Jung, T. F. Waddell, and K. J. Kim, "Cheery companions or serious assistants? Role and demeanor congruity as predictors of robot attraction and use intentions among senior citizens," *Int. J. Hum.-Comput. Stud.*, vol. 97, pp. 88–97, Jan. 2017.

[39] A. Vega, K. Ramírez-Benavides, L. A. Guerrero, and G. López, "Evaluating the nao robot in the role of personal assistant: The effect of gender in robot performance evaluation," *Multidisciplinary Digit. Publishing Inst. Proc.*, vol. 31, no. 1, p. 20, 2019.

[40] M. J. A. Craig and C. Edwards, "Feeling for our robot overlords: Perceptions of emotionally expressive social robots in initial interactions," *Commun. Stud.*, vol. 72, no. 2, pp. 251–265, Mar. 2021.

[41] P. Dugard and J. Todman, "Analysis of pre-test-post-test control group designs in educational research," *Educ. Psychol.*, vol. 15, no. 2, pp. 181–198, Jan. 1995.

[42] A. Salim, A. Mackinnon, H. Christensen, and K. Griffiths, "Comparison of data analysis strategies for intent-to-treat analysis in pre-test-post-test designs with substantial dropout rates," *Psychiatry Res.*, vol. 160, no. 3, pp. 335–345, Sep. 2008.

[43] K. L. Koay, D. S. Syrdal, M. L. Walters, and K. Dautenhahn, "Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human–robot interaction study," in *Proc. 16th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, vol. 5, Aug. 2007, pp. 564–569.

[44] S. Ikemoto, H. B. Amor, T. Minato, B. Jung, and H. Ishiguro, "Physical human–robot interaction: Mutual learning and adaptation," *IEEE Robot. Autom. Mag.*, vol. 19, no. 4, pp. 24–35, Dec. 2012.

[45] C. H. Rankin, T. Abrams, R. J. Barry, S. Bhatnagar, D. F. Clayton, J. Colombo, G. Coppola, M. A. Geyer, D. L. Glanzman, S. Marsland, F. K. McSweeney, D. A. Wilson, C.-F. Wu, and R. F. Thompson, "Habituation revisited: An updated and revised description of the behavioral characteristics of habituation," *Neurobiol. Learn. Memory*, vol. 92, no. 2, pp. 135–138, Sep. 2009.

[46] K. L. Koay, D. S. Syrdal, M. L. Walters, and K. Dautenhahn, "Five weeks in the robot house—Exploratory human–robot interaction trials in a domestic setting," in *Proc. 2nd Int. Conf. Adv. Comput.-Hum. Interact.*, vol. 7, Feb. 2009, pp. 219–226.

[47] K. Bergmann, F. Eyssel, and S. Kopp, "A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time," in *Proc. Int. Conf. Intell. Virtual agents*. Cham, Switzerland: Springer, 2012, pp. 126–138.

[48] M. Giuliani, R. P. A. Petrick, M. E. Foster, A. Gaschler, A. Isard, M. Pateraki, and M. Sigalas, "Comparing task-based and socially intelligent behaviour in a robot bartender," in *Proc. 15th ACM Int. Conf. Multimodal Interact.*, Dec. 2013, pp. 263–270.

[49] E. T. van Dijk, E. Torta, and R. H. Cuijpers, "Effects of eye contact and iconic gestures on message retention in human–robot interaction," *Int. J. Social Robot.*, vol. 5, no. 4, pp. 491–501, Nov. 2013.

[50] K. S. Haring, Y. Matsumoto, and K. Watanabe, "Perception and trust towards a lifelike Android robot in Japan," in *Proc. Trans. Eng. Technol., Special Issue World Congr. Eng. Comput. Sci.*, 2014, pp. 485–497.

[51] S. Keizer, P. Kastoris, M. E. Foster, A. Deshmukh, and O. Lemon, "Evaluating a social multi-user interaction model using a nao robot," in *Proc. 23rd IEEE Int. Symp. Robot Hum. Interact. Commun.*, Aug. 2014, pp. 318–322.

[52] S. Keizer, M. E. Foster, A. Gaschler, M. Giuliani, A. Isard, and O. Lemon, "Handling uncertain input in multi-user human–robot interaction," in *Proc. 23rd IEEE Int. Symp. Robot Hum. Interact. Commun.*, Aug. 2014, pp. 312–317.

[53] R. H. Cuijpers and M. A. Knops, "Motions of robots matter! The social effects of idle and meaningful motions," in *Proc. 7th Int. Conf. Social Robot. (ICSR)*. Cham, Switzerland: Springer, 2015, pp. 174–183.

[54] K. S. Haring, K. Watanabe, D. Silvera-Tawil, M. Velonaki, and T. Takahashi, "Changes in perception of a small humanoid robot," in *Proc. 6th Int. Conf. Autom., Robot. Appl. (ICARA)*, Feb. 2015, pp. 83–89.

[55] M. K. X. J. Pan, E. A. Croft, and G. Niemeyer, "Evaluating social perception of human-to-robot handovers using the robot social attributes scale (RoSAS)," in *Proc. 13th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2018, pp. 443–451.

[56] M. Paetzel and G. Castellano, "Let me get to know you better: Can interactions help to overcome uncanny feelings?" in *Proc. 7th Int. Conf. Hum.-Agent Interact.*, vol. 6, Sep. 2019, pp. 59–67.

[57] R. Paradeda, M. J. Ferreira, R. Oliveira, C. Martinho, and A. Paiva, "What makes a good robotic advisor? The role of assertiveness in human–robot interaction," in *Proc. 11th Int. Conf. Social Robot. (ICSR)*. Cham, Switzerland: Springer, 2019, pp. 144–154.

[58] J. Rhim, A. Cheung, D. Pham, S. Bae, Z. Zhang, T. Townsend, and A. Lim, "Investigating positive psychology principles in affective robotics," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2019, pp. 1–7.

[59] A. Ueno, K. Hayashi, and I. Mizuuchi, "Impression change on nonverbal non-humanoid robot by interaction with humanoid robot," in *Proc. 28th IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Oct. 2019, pp. 1–6.

[60] M. Paetzel, G. Perugia, and G. Castellano, "The persistence of first impressions: The effect of repeated interactions on the perception of a social robot," in *Proc. 15th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2020, pp. 73–82.

[61] Y. Terzioglu, B. Mutlu, and E. Sahin, "Designing social cues for collaborative robots: The RoIe of gaze and breathing in human–robot collaboration," in *Proc. 15th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2020, pp. 343–357.

[62] Y. Xiao, D. Silvera-Tawil, and M. Pagnucco, "Autonomous behaviour planning for socially-assistive robots in therapy and education," in *Proc. Australas. Conf. Robot. Autom.*, 2020, pp. 1–9.

[63] I. P. Bodala, N. Churamani, and H. Gunes, "Teleoperated robot coaching for mindfulness training: A longitudinal study," in *Proc. 30th IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2021, pp. 939–944.

[64] G. Perugia, M. Paetzel-Prüsmann, M. Alanenpää, and G. Castellano, "I can see it in your eyes: Gaze as an implicit cue of uncanniness and task performance in repeated interactions with robots," *Frontiers Robot. AI*, vol. 8, pp. 1–18, Apr. 2021.

[65] P. K. John, X. Zhu, T. Gedeon, and W. Zhu, "Evaluating human impressions of an initiative-taking robot," in *Proc. CHI Conf. Hum. Factors Comput. Syst. Extended Abstr.*, vol. 49, Apr. 2022, pp. 1–7.

[66] A. Pipitone, A. Geraci, A. D'Amico, V. Seidita, and A. Chella, "Robot's inner speech effects on human trust and anthropomorphism," *Int. J. Social Robot.*, vol. 16, pp. 1333–1345, Jul. 2023.

[67] M. Spitale, M. Axelsson, and H. Gunes, "Robotic mental well-being coaches for the workplace: An in-the-wild study on form," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, vol. 17, Mar. 2023, pp. 301–310.

[68] J.-P. Stein and J. Banks, "Valenced media effects on robot-related attitudes and mental models: A parasocial contact approach," *Hum.-Mach. Commun.*, vol. 6, pp. 155–182, Jul. 2023.

[69] I. Tusseyeva, A. Oleinikov, A. Sandygulova, and M. Rubagotti, "Evaluation of perceived intelligence for a collaborative manipulator sharing its workspace with a human operator," in *Proc. 32nd IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2023, pp. 2267–2272.

[70] P. Dautzenberg, S. Ladwig, and A. M. Rosenthal-von der Pütten, "Follow me: Anthropomorphic appearance and communication impact social perception and joint navigation behavior," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, Mar. 2024, pp. 175–183.

[71] R. M. Warner and D. B. Sugarman, ''Attributions of personality based on physical appearance, speech, and handwriting,'' *J. Personality Social Psychol.*, vol. 50, no. 4, pp. 792–799, 1986.
[72] C. Bartneck, M. Verbunt, O. Mubin, and A. A. Mahmud, ''To kill a mockingbird robot,'' in *Proc. 2nd ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2007, pp. 81–87.
[73] C. Bartneck, M. van der Hoek, O. Mubin, and A. A. Mahmud, '''Daisy, daisy, give me your answer do!' Switching off a robot,'' in *Proc. 2nd ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2007, pp. 217–222.
[74] C. Bartneck and J. Hu, ''Exploring the abuse of robots,'' *Interact. Stud. Social Behav. Commun. Biol. Artif. Syst.*, vol. 9, no. 3, pp. 415–433, Dec. 2008.
[75] S. Thunberg, M. Arnelid, and T. Ziemke, ''Older adults' perception of the furhat robot,'' in *Proc. 10th Int. Conf. Hum.-Agent Interact.*, Dec. 2022, pp. 4–12.
[76] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, ''The robotic social attributes scale (RoSAS): Development and validation,'' in *Proc. 12th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI*, Mar. 2017, pp. 254–262.
[77] S. X. Liu, Q. Shen, and J. Hancock, ''Can a social robot be too warm or too competent? Older Chinese adults' perceptions of social robots and vulnerabilities,'' *Comput. Hum. Behav.*, vol. 125, pp. 1–7, Aug. 2021.
[78] L. Riek, ''Wizard of oz studies in HRI: A systematic review and new reporting guidelines,'' *J. Hum.-Robot Interact.*, vol. 1, no. 1, pp. 119–136, Aug. 2012.
[79] M. Mori, K. F. MacDorman, and N. Kageki, ''The uncanny valley [from the field],'' *IEEE Robot. Autom. Mag.*, vol. 19, no. 2, pp. 98–100, Jun. 2012.
[80] A. Müller, M. Schiffmann, A. Neumeister, and A. Richert, *Exploring Beyond the Exhibits: Creating Knowledge for Social Robots in Public Spaces*. AI in Museums, 2023, pp. 273–286.
[81] E. Mingotto, F. Montaguti, and M. Tamma, ''Challenges in re-designing operations and jobs to embody AI and robotics in services. Findings from a case in the hospitality industry,'' *Electron. Markets*, vol. 31, no. 3, pp. 493–510, Sep. 2021.
[82] L. Blavette, A.-S. Rigaud, S. M. Anzalone, C. Kergueris, B. Isabet, S. Dacunha, and M. Pino, ''A robot-mediated activity using the nao robot to promote COVID-19 precautionary measures among older adults in geriatric facilities,'' *Int. J. Environ. Res. Public Health*, vol. 19, no. 9, p. 5222, Apr. 2022.

INARA TUSSEYEVA received the master's degree in computer science from Gyeongsang National University, Jinju, South Korea, in 2013, and the Ph.D. degree in robotics engineering from Nazarbayev University, Astana, Kazakhstan, in 2024. She was with the Engineering Technical Center, Management Department of the President of Kazakhstan, as a Software Developer, and currently holds the position of an Assistant Professor with the Department of Intelligent Systems and Cybersecurity, Astana IT University, and a Research Assistant with the Institute of Smart Systems and Artificial Intelligence, Nazarbayev University. Her research interests include robot motion planning and physical human–robot interaction, with particular focus on the assessment of perceived safety and perceived intelligence.

ANARA SANDYGULOVA received the Ph.D. degree in computer science from University College Dublin, Ireland. She holds the position of an Associate Professor with the Department of Robotics and Mechatronics, Nazarbayev University (NU). Her expertise and experience in the domain of human–robot interaction research and her primary focus centers on research and development of sign language processing, robot-assisted learning, and therapy solutions. She is the Director of the Human–Robot Interaction (HRI) Laboratory, NU, which is dedicated to interdisciplinary research integrating robotics, artificial intelligence, psychology, education, and child development. She has published more than 80 research articles.

MATTEO RUBAGOTTI (Senior Member, IEEE) received the Ph.D. degree in electronics, computer science, and electrical engineering from the University of Pavia, Pavia, Italy, in 2010. He held a postdoctoral positions with the University of Trento, Trento, Italy, and the IMT Institute for Advanced Studies, Lucca, Italy; and a faculty positions with the University of Leicester, Leicester, U.K., and Nazarbayev University, Astana, Kazakhstan, where he is currently a Professor of robotics. His current research interests include model predictive control, sliding mode control and imitation/reinforcement learning, and their applications to robotics and mechatronics. He is the Subject Editor of the *International Journal of Robust and Nonlinear Control* and a member of the conference editorial boards of the IEEE Control System Society and of European Control Association.