International Journal of Social Robotics (2023) 15:473–500
https://doi.org/10.1007/s12369-021-00785-7

Communication Models in Human–Robot Interaction: An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI)

Helena Anna Frijns[1] · Oliver Schürer[2] · Sabine Theresia Koeszegi[1]

Accepted: 10 April 2021 / Published online: 3 May 2021

Abstract

We argue for an interdisciplinary approach that connects existing models and theories in Human–Robot Interaction (HRI) to traditions in communication theory. In this article, we review existing models of interpersonal communication and interaction models that have been applied and developed in the contexts of HRI and social robotics. We argue that often, symmetric models are proposed in which the human and robot agents are depicted as having similar ways of functioning (similar capabilities, components, processes). However, we argue that models of human–robot interaction or communication should be asymmetric instead. We propose an asymmetric interaction model called AMODAL-HRI (an Asymmetric MODel of ALterity in Human–Robot Interaction). This model is based on theory on joint action, common robot architectures and cognitive architectures, and Kincaid's model of communication. On the basis of this model, we discuss key differences between humans and robots that influence human expectations regarding interacting with robots, and identify design implications.

Keywords Human–Robot interaction · Communication theory · Communication models · Interaction design · Human–Machine communication

1 Introduction

In the Human–Robot Interaction (HRI) literature, models of interpersonal communication have been applied (see for instance [59,93]), and several models for human–robot interaction and communication have been developed (e.g. [44,52,91]). Concepts from communication theory have been discussed and applied, but often with little theoretical context. The present paper aims to fill this gap through thorough reflection on existing models in the literature on communication between humans and HRI, in order to (1) connect existing models and theories in the field of Human–Robot Interaction to different traditions within communication theory, (2) critically discuss (symmetric) models of HRI and (3) formalize an asymmetric model for human–robot joint action. Our main aim is to make the asymmetries between a human and a robot agent that are engaged in a communication process explicit, in order to provide design guidelines that mitigate potential communication failures arising from these asymmetries.

The first aim of this article is to connect communication theory with HRI and social robotics, an interdisciplinary endeavour. Researchers in HRI already apply concepts from semiotics [43], for instance, but the connection to a broader research field and theory of what constitutes communication is often lacking. We aim to go beyond simply borrowing concepts and models from communication research, and instead connect these concepts and theories to the broader field of communication theory. This will hopefully give HRI researchers a better overview of entry points and existing theory on models of human communication. We also wish to point to the potential of communication theory as a practical discipline that can serve to inform the design of robotic systems.

The second aim is to critically discuss communication models as currently applied in HRI. We identify shortcomings of existing models. We posit that current communication

✉ Helena Anna Frijns
helena.frijns@tuwien.ac.at

Oliver Schürer
schuerer@attp.tuwien.ac.at

Sabine Theresia Koeszegi
sabine.koeszegi@tuwien.ac.at

[1] TU Wien, Institute of Management Science, Theresianumgasse 27, Vienna, Austria

[2] TU Wien, Department of Architecture Theory, Karlsplatz 13, Vienna, Austria

models of HRI are lacking with respect to the context dependency of interactions, the influence of external actors, and asymmetry between humans and robots. With asymmetry, we mean that robots and humans are at present fundamentally different entities, and rather than focusing solely on their (theoretical) similarities in models and designs, we should carefully consider their differences. Models for communication between two agents that are 'symmetrical' (reflectional or bilateral symmetry) presuppose that we can use the exact same components to model any agent and that both agents have similar requirements and ways of functioning within the interaction. An asymmetric model, on the other hand, does not assume this. We argue that acknowledging differences between humans and robots should be embedded in models and robot designs. Other scholars have similarly argued that human–robot interaction should be conceived as asymmetric [56,77], and emphasize that there are functional, physical and cognitive differences between humans and robots [21]. Guzman and Lewis [41] argue that the similarities and differences between humans and robots need to be assessed. When it comes to modelling human–robot communication, other researchers have argued that a new model of human–robot communication is required: a model including facets of communication that remain implicit in existing models (for example, knowledge of mission goals and cultural norms [93]). While the concept of asymmetry has previously been proposed by other researchers [42,56,76,77], the consequences of this concept have not been analysed in detail with regards to communication modelling and interaction design, which is what we set out to do here. See Section 2.3 for a more detailed discussion of the concept of asymmetry.

Building upon this second aim, we propose a model for human–robot joint action and communication, our third aim. We use this model to discuss the differences between the human and the robot side of the model, and highlight design implications based on the model and the identified differences. We will highlight how this model is asymmetric and how both the human and the robot agent can be influenced by external actors. We argue that such a model contributes to an enhanced understanding of key differences between humans and robots, and how these differences can conflict with human expectations of the interaction. This, in turn, can help us reconsider robot design (behaviour and embodiment) and increase usability. The model is called AMODAL-HRI (Asymmetric MODel of ALterity in Human–Robot Interaction). The name 'AMODAL' is deemed fitting, as amodal completion (or *"[t]he perception of complete objects behind occluders"* [29, p.1188]) refers to the phenomenon of perceiving an object as whole even if it is only partially perceivable. A similar phenomenon can occur with respect to robots: people may perceive a robot as a being with agency, even though it is only made up of a collection of technical components. By introducing a "model of alterity", we uncover and dismantle the differences between human and robot actors, which allows for making the best of use of their complementary capabilities. *Alterity*, or otherness, refers to the term alterity as developed in phenomenology. The phenomenologist Don Ihde uses the term *alterity* to describe a particular set of human-technology relations, specific to relating to *"technology-as-other"*. The *quasi-otherness* that Ihde describes, indicates that some technological artefacts occupy a status between objects and human or animal otherness [50]. This understanding of certain technologies as *quasi-other* leads to the understanding of relations to those technologies as alterity relations [51]. We emphasize that this is the sense of 'otherness' referred to in the model name, not a human otherness. Otherness applied to the technological artefact refers to the fundamental otherness of the robot as a sociotechnical assemblage that is subject to outside control. See also Sect. 2.3.

We argue that connecting communication theory on human–human interaction to HRI and highlighting asymmetries is important, firstly because a robot is an embodied entity that acts (to some extent) autonomously in physical space, with actions that are communicative to a human interaction partner (and those actions can be expressive of, or be interpreted to be indicative of, agency), and secondly because of the anthropomorphic design strategies that are employed in social robotics and HRI. We argue that there are both advantages and disadvantages to modelling robot communicative behaviours based on human ones. The advantages are that theories on communication and interaction between humans give us significant insight into what humans might expect from a robotic interaction partner and what is necessary for their collaboration to be successful. Krämer et al. [58] argue that there is no real alternative to using theory from human–human interaction, as humans will expect communicative mechanisms similar to those they are familiar with from interactions with other humans, though we should ensure that the theory we use is applicable to a device or robot context. Potential disadvantages are (1) that human expectations of a robotic interaction partner may be too high if humanlike communication behaviours are implemented on the robot [7], and (2) that a focus on humanlike embodiments and behaviours may be restrictive (using alternative design strategies may lead to more diverse, more successful designs) [76]. Therefore, we argue that acknowledging differences between humans and robots rather than pursuing similarity alone is a relevant additional design strategy in order to (1) avoid communication failures when a human interacts with a robot and expects human-level functioning, and (2) expand the range of possibilities and make it explicit that we do not necessarily need to model robot bodies and behaviours on human ones.

The structure of this article is as follows. First, in Sect. 2, we discuss definitions of communication and interaction as well as key research traditions in communication theory. We will also explain what we mean with asymmetry in more detail. In Sect. 3, we discuss general communication models that have been developed to model interpersonal communication and their application to HRI. In Sect. 4, we discuss models that have been specifically developed in the context of HRI and the different types of interaction they represent. In Sect. 5, we discuss different ways interaction and communication are conceptualized in HRI. In Sect. 6, we propose an asymmetric model for HRI (AMODAL-HRI) that is intended to be used for comparing human and robot agents, and identifying differences in capabilities that can lead to problems regarding human expectations of the interaction. We identify differences between the human and the robot agent in our model and give several design recommendations based on those differences in Sect. 7.

Note that our discussion will focus on models of communication and their application to HRI and social robotics. Additional topics, such as sociality, language, signs and signals are discussed as important aspects within models of communication that have received attention in the context of robotics, but are not the main focus of this article.

2 Interaction and Communication: What Do These Terms Mean in the Context of HRI?

In this section, we first provide some theoretical background from the field of communication theory. In Sect. 2.2, we discuss the definition of interaction between one or more humans and one or more robots. We also discuss whether the term 'communication' applies to interactions between humans and robots. In Sect. 2.3, we explain the concept of asymmetry in HRI.

2.1 What is 'Communication'?

In this section, we establish that there are several different traditions in communication theory, and that each tradition views communication in different ways. Definitions of communication differ in their level of abstraction, whether they describe communication as intentional (having a particular goal) and whether the definition includes a normative evaluation (for instance, effectiveness) [63]. A multitude of definitions are possible, depending on the goal of the person who proposes the definition. When we talk about communication, we can discuss this topic at different levels of detail. Littlejohn and Foss distinguish between the level of the communicator, the message, the conversation, the relationship, groups and organizations, the media, and finally society and culture, at increasing orders of magnitude and complexity [63]. Guzman and Lewis [41], whose work is situated in the research area of Human–Machine Communication, argue that interaction with AI departs from traditional communication theory, as AI technologies have begun to take on the role of the communicator, a role that, in communication theory, could previously only be performed by humans. We discuss communication at different levels throughout this paper, but focus on the levels of the communicator, the message and the conversation.

Craig proposed the constitutive theory or constitutive metamodel of communication [20]. This model does not directly describe the communication process (it is not a first-order model), but is rather a second-order model that incorporates different views and traditions in communication research. Craig proposed that several traditions can be distinguished within the broad research field of communication theory. Communication had been an object of study within many domains, but no clear discipline had emerged by the time Craig wrote his article. The constitutive metamodel can be understood as an attempt to frame communication theory as useful as a metadiscourse (discourse regarding discourse) and as a practical discipline, oriented to the discussion of practical, real-world phenomena [20]. The traditions that Craig distinguishes are the rhetorical, cybernetic, semiotic, phenomenological, sociopsychological, sociocultural, and critical traditions in communication theory [63]. These traditions are not to be seen as incompatible or completely separate; combinations and overlaps are common. In each tradition, communication is understood in a different way:

- Rhetorical tradition: communication as a *"practical art of discourse"* [20, p.135]
- Semiotic tradition: communication as *"intersubjective mediation by signs"* [20, p.136]
- Phenomenological tradition: communication as *"dialogue"* or the *"experience of otherness"* [20, p.138]
- Cybernetic tradition: communication as *"information processing"* [20, p.140]
- Sociopsychological tradition: communication as *"a process of expression, interaction, and influence"* [20, p.143]
- Sociocultural tradition: communication as *"a symbolic process that produces and reproduces shared sociocultural patterns"* [20, p.144]
- Critical tradition: communication as *"discursive reflection"* [20, p.147]

Some of these traditions may be more useful to HRI researchers than others; however, it should be kept in mind that one tradition by itself will not suffice for describing all that is relevant to communication. With regard to HRI, we can distinguish lines of research applying views of communication that can be linked to the traditions identified by Craig. The views that are common in HRI are mostly in line with the semiotic, cybernetic, sociopsychological and sociocultural traditions[1].

In the semiotic tradition, the capability to use (human) language[2] is of particularly high importance. However, we can also speak of communication in other animal species [62], and animals such as dolphins have been shown to be capable of at least some aspects of language comprehension and production [70]. Besides the capacity to use verbal or written language or other symbol systems, we consider nonverbal behaviour, body language, and other ways of signalling to be part of communicating. Nonverbal behaviour includes gestures and body movements (including, for instance, eye gaze), proxemics, touch, and appearance [86].

2.2 Can We Speak of 'Communication' with Respect to Interactions Between Humans and Robots?

Before discussing human–robot communication specifically, how might we define *interactions* between humans and robots? In the context of HRI, Bensch et al. propose a model that treats interaction as *"an interplay between human(s), robot(s), and environment"* [4, p. 184]. Goodrich and Schultz describe interaction in the context of HRI as *"the process of working together to accomplish a goal"* [39, p. 217]. They propose the concept of dynamic interaction as a characterization that incorporates all five dimensions HRI designers can affect, namely autonomy, how information is exchanged, team structure, learning and training of the humans and robots involved, and the shape of the task [39].

How does communication relate to interaction?

Goodrich and Schultz [39] take communication as a requirement for human–robot interaction; in their view, interaction is, in some form, present in every robot application. They separate both communication and interaction in HRI into the categories proximate and remote interaction. Based on their text, we infer that their view of communication is that of exchanging information. Their description of information exchange in HRI focuses on the media used in communication processes and the format of communication. Visual, auditory and tactile modalities are most relevant in HRI, and these are typically present in forms such as: visual displays, gestures, natural language (text and spoken language), audio, physical interaction, and haptics [39].

These interaction modalities can also be combined in multimodal interfaces, enabling the user to interact with the system by means of multiple communication modes, and providing the user with multimedia output. Advantages of multimodal interaction are that they can better support users' preferences, enhance the expressive power of the user interface, reduce user errors, and lead to small efficiency improvements [27]. Other reasons to use multimodal interfaces can include reducing the human's cognitive workload and increasing the ease of learning to use the interface [39].

The user interface is what allows the human to interact with the robot, or allows the human to interact with an environment using the robot. The concept of the user interface (as a restricted area reserved for information exchange) is problematized in robotics for co-located robots, as in this case we no longer have a restricted area that is reserved for interaction (such as a screen). Instead, the entire embodiment of the robot, which acts (semi-)autonomously in its environment, becomes communicative or informative to the user. In remote interaction, as found in teleoperation applications, by contrast, the user can access the robot and its environment by means of a screen and control modalities, which is more similar to traditional device operation. Here, the user interface is restricted to devices that allow the user to exchange information with the robot and to affect both the robot and its environment.

Should we use the word "communication" to describe the interaction between human(s) and robot(s)? According to Seibt, using intentionalist vocabulary to describe robots is confusing, inaccurate and imprecise. She proposes OASIS, the Ontology of Asymmetric Social Interactions, which provides a description language for further theory developments, and offers the possibility to include non-humans as social interaction partners while emphasizing differences with human social interaction. Seibt argues that in order to "work with" robots, for instance, robots would need to be capable of having the phenomenological experience of working [77]. Instead of arguing that a robot is capable of communication, we can apply Seibt's framework and say that the robot is either functionally replicating, imitating, mimicking, displaying or approximating the process of communicating. In the context of this article, however, we are mostly concerned with the human–robot system and the way

[1] For a discussion of how semiotics is studied in the context of HRI, see Sect. 5.1. Theories that fall within the sociopsychological tradition of communication theory, such as communication accommodation theory and interaction adaptation theory [63], have been studied in the context of HRI as well, see interactional synchrony for example [65]. For work in HRI in the sociocultural vein, see Sect. 4.2. Sandry applied traditions in communication theory to different types of robots in order to analyse them. For instance, she analyses the robot Kismet using the sociopsychological and sociocultural traditions [76].

[2] Language has been described as *"a multifaceted and complex ability that allows us to assign arbitrary symbols meaning and to use and understand these symbols in referential exchanges with others that draw joint attention to agents, objects, and events both present and displaced in space and time"* [70, p.1]. Lindesmith et al. describe that language has the following properties: language consists of combinations of meaningless sounds into words that are meaningful (duality) and new (productivity). The relation between symbol and signified is arbitrary (arbitrariness). The speaker has the possibility to discuss things that are not in the same time and location as the speaker (displacement), and capacity for cultural transmission (learning instead of genetic transmission of information). Words have distinct functions (specialization) and speakers can recreate and reproduce messages (interchangeability) [62].

the robot appears to the human (as opposed to what the robot is capable of by itself). We would argue that to a human interaction partner, the robot will come across as communicating, at the very least in the sense of exchanging information. Although it may be imprecise to speak of "communicating", we maintain the term for sake of practicality. However, we urge the reader to keep Seibt's proposal in mind. While it can be argued that there is no communication with the robot but instead with its designers or developers, this is not a sufficient explanation for communication or interaction with (semi-)autonomous systems. In such a case, there is an element of unpredictability in the interaction and how the human and the robot relate to the current situation. In such a case, the human could be said to be communicating with a dynamic system consisting of a robot and its designers, developers, maintainers, et cetera, with the communication process focused on or enabled by the embodied robot. For our purposes, we define communication in the context of HRI as actions performed by human and robotic agents that have aims such as coordinating behaviour, reducing uncertainty, and building a common understanding.

2.3 Asymmetry

Coeckelbergh [17] posits that human–robot relations can be described as social relations, since robots perform roles in society and participate in interactions with humans that can be described as quasi-social. He describes robots with the term *quasi-others* to indicate their appearance as social actors. Alaç [1] proposes that in certain settings, both thing-like and agent-like characteristics of social robots are present. These different characteristics surface at different moments in the interaction.

While humans can experience a robot as a social entity, this does not take away from the fact that humans and robots are very different, and that robots are at present very limited in their abilities. People may *experience* robots as social entities. However, robots do not have the same capabilities or responses as humans with respect to social interaction. Therefore, we should in a theoretical discussion highlight these differences and study how they can become relevant for the design of robots and robot behaviours. This can also help identify when expectations on the human side may arise regarding social responses by the robot.

Generally, it can be said that any two agents (humans or otherwise) who engage in a communication process are different from each other to some extent. They bring different sets of background knowledge to the interaction, different needs, different bodies, different (cognitive) abilities, etc. We can describe this as a kind of 'hidden' asymmetry that needs to be made (more) explicit when discussing communication and interaction at a high level of abstraction. However, in human–human interaction, it can also generally be assumed that there is a significant level of similarity between the interaction partners. We can assume some level of shared background knowledge when interacting with another person, we can expect that our interaction partner will abide by similar conventions, will often speak a language we are familiar with, and if not, at least have similar needs such as the need for food and water, et cetera.

The situation is different for human–robot interaction, however. In this context, we cannot assume similar processing mechanisms, similar background knowledge, or similar functional, cognitive or physical capabilities. Instead, human–robot interaction and communication are better described as asymmetric.

Other authors have made similar arguments, most notably with regards to agency (which refers to the capability to act in an autonomous way [98]). Seibt [77] argues that interactions between humans and robots *"form a new type of social interaction("asymmetric social interactions") where the capacities for normative agency are not symmetrically distributed amongst interaction partners and which therefore are not by default potentially reciprocal, as this is the case with social interactions among humans"* [77, p.135]. Kruijff [56, p.154] has argued that robots are functionally asymmetric to humans and has proposed the concept of asymmetric agency. This concept refers to a group of agents in which individual members of the group have different understandings of reality ("asymmetry in understanding"), which in turn can result in different expectations regarding ways of acting in this (differently-understood) reality. In addition to asymmetry in understanding and capacities to act, symmetry and asymmetry between humans and robots can also be identified on the level of embodiment. Sandry [76] argues that the development of humanoid robots and of human-like communication mechanisms for robots point to a pursuit of commonality and a view of communication as the transmission of information. This could be described as using 'sameness' as a design strategy, which she is critical of. Instead of striving for "complete comprehension", she argues for striving for a "partial understanding" that recognizes the *"alterity of the machine"* [76, p.8]. Hassenzahl et al. note that humans perceive robotic and AI systems as counterparts instead of tools and write *"it creates a fundamental shift from an* embodied *relationship with technology to one of* alterity: *Technology becomes other"* [42, p.54]. They call such systems *otherware*. Mimicking human or animal communication strategies comes at the risk of stereotypical designs and disappointment. They propose *animistic design* as an alternative design strategy, and point to a need for the HCI community to develop new models, interaction paradigms, design patterns and design methods for otherware [42]. Gunkel notes that *"Communication studies (…) must (…) reorient its theoretical framework so as to be able to accommodate and respond to situations where the* other *in communicative exchange is no longer exclusively human"* [40, p.2].

Sandry [76] argues that humans and robots are different entities and that this otherness of robots can be valuable, instead of a problem to be overcome. She argues that this is made difficult by the fact that in communication theory, communication is often framed as accurate information exchange or a means to reproduce shared social patterns. In other words, communication is often viewed as a means to emphasize what communicators have in common, and increase the similarity between those who are communicating. However, this comes at a loss, as the *other* has different points of view that are devalued in a communication process which is aimed at enhancing similarity. Based on work by Pinchevski, she notes that such an understanding of communication can even be described as *"violent to the other"* [76, p.5]. While this is a far greater issue in human–human communication, the disadvantage of not acknowledging differences is a potential loss of possibilities [76]. In addition, a culture of emphasizing what humans and robots have in common comes at the risk of rendering humans as computational. For HRI, this could result in missed opportunities to design other behaviours and embodiments that are not based on the human model. In human–robot teams, not acknowledging differences poses a risk for team functioning; we should acknowledge that capacities are different across the members of a human–robot team and strive to make the most of complementary capabilities. Johnson et al. propose a method of interdependence analysis to analyse the capacities of different team members and how they depend on one another, in order to make use of complementary capabilities [52]. Sandry argues that it is the difference, the complementarity of humans' and robots' skills and the coordination of these skill sets that makes collaboration successful. While humans take responsibility and usually instructs other team members, robots have other important roles to play [76].

To summarize, we view human–robot interactions as asymmetric, as the interaction partners function in different ways due to differences in embodiment, cognitive capabilities, functional capabilities, and capacities for social interaction. We expect this asymmetry to remain in place in the foreseeable future, even with large improvements to robot (cognitive) functioning. Like Sandry, we view asymmetry between human and robot interaction partners as a potentially productive, useful feature. However, we argue if this asymmetry is not acknowledged (for instance, by striving to make humans and robots function as similarly as possible, or by ignoring the existence of asymmetry), this can be problematic and result in communication failures. Assuming symmetry where there is none can be productive for initial engineering attempts, but fail to identify problems in interactions with humans. Asymmetric models are more suitable tools that can foresee at least some of these problems. Therefore, we propose an asymmetric model of Human–Robot social interactions in Sect. 6 and design recommendations based on the identified asymmetries in Sect. 7.

3 Classical Models of Communication and Interaction

In this section we discuss existing models of communication between humans. In Sect. 2.1, we already introduced Craig's constitutive metamodel [20]. One can distinguish several different types of communication models, with two well-known types being transmission and transactional models. Authors have discussed communication in different ways, depending on their goals. For HRI, a key challenge is to establish shared awareness of a team task and to coordinate actions to achieve the task goals, which is why transactional models remain relevant in this context. Other types of models that take a different perspective on communication (e.g. with a focus on power relations) can also be insightful.

3.1 The Transmission Model of Communication

In transmission models of communication (or also: linear or container models of communication), communication is described as the one-way transmission of a message from a source to a receiver. These types of models serve to depict the way technological communication functions, and are used to study the process of making sure a signal arrives at its destination intact so that the original message can be reconstructed [78]. In such models, feedback and context are not considered. The message is viewed as a kind of container for meaning that is transferred from A to B (thus following a postal metaphor [8]). The most well-known transmission model of communication is outlined in the article *The Mathematical Theory of Communication* by Shannon and further developed by Weaver. Their focus was on communication systems such as telegraph, radio, and telephone systems. The model consists of a chain in which information moves from the information source, to the transmitter (which sends a signal over a channel), to the receiver, which reconstructs the message and sends it to the final destination. The message can be corrupted by a source of noise [78]. This model falls within the cybernetic tradition in communication theory.

Transmission models have been criticized for being linear and one-way [54]. Such models exhibit epistemological biases in that they treat information like a physical substance that can be carried from point A to point B, and treat minds as disembodied entities, stripped of their context. Additionally, Kincaid argues that such models focus on communication as a means of persuasion and focus on individual psychological effects rather than effects on the social whole and social relationships. A one-way model implies one-way causation;

there is no space for mutual causation [54]. The signal may be corrupted by noise, but otherwise the signal should stay the same until it is decoded by the receiver. The model is useful for its purpose, which is to describe the technological process of sending a message from A to B, but not sufficient for modelling how shared meaning arises in interpersonal communication[3]. Even though the transmission model of communication has been criticized, it has frequently reappeared in the HRI literature [44,67], often with feedback loops added to turn it into a transactional model of communication.

3.2 Transactional Models of Communication

Transactional models of communication introduce the possibility for feedback from the receiver to the sender; they depict humans involved in communication as both senders and receivers. Additionally, such models often include contextual factors that influence communication. With respect to these models, communication can be described as having the goal of arriving at mutual understanding [54] or building shared meaning and reducing uncertainty [2]. Such models often aim to identify relevant components or factors that influence human communication rather than describing the technical process of communication.

Barnlund [2] describes communication as dynamic, continuous, circular, unrepeatable, irreversible, and complex. People involved in communication *construct* meaning on the basis of the other person's messages, rather than reconstruct it. This construction of meaning should assist in deciding on a course of action that is likely to be effective and fits the demands of the current situation. Barnlund proposes 'pilot models' of a transactional model of communication. Barnlund discusses that there can be limitless cues involved in a transactional communication process (public, private, natural, artificial, behavioural verbal and behavioural nonverbal cues). Kincaid [54] applies perspectives from cybernetics to propose the convergence model of communication. This model views communication as a process. The aim of arriving at mutual understanding is achieved by creating and sharing information, which the participants in the communication process interpret. While they may converge on meaning and therefore increase their mutual understanding, they can never converge completely because each individual brings their own set of experiences to the communication process. Communication occurs within humans' psychological, physical and social realities, and building mutual understanding is supported by the actions and beliefs of both parties [54]. Because this model describes the goal of communication as reaching mutual understanding, it can also be understood as a transactional model. We will use this model later as the foundation for a model of human–robot interactions from a joint action perspective.

Classical transactional models of communications have also been discussed in the context of HRI, including the circular model of communication as proposed by Osgood and Schramm (which depicts agents as processing a message via a decoder, interpreter and encoder, and the messages between them being sent along a continuous loop, depicted by arrows between the entities) and Berlo's model of communication (with the main components source, message, channel, and receiver) [93]. Lackey et al. [59] discuss the transactional model by Barnlund in the context of HRI.

Pickering and Garrod [72] criticize transactional models of language processing that explicitly separate production and comprehension processes. Instead, they propose that production and comprehension processes need to be tightly interwoven to support agents in coordinating their actions, resulting in joint action. Agents not only predict their own actions, they also predict the actions of the agent they are interacting with. Thus, the actions of both agents can become tightly coupled [72].

4 Communication and Interaction Models in HRI

As discussed in the previous section, models describing interpersonal communication have also been applied to HRI. In this section, we review models that have been developed specifically for HRI. Some of these models are based on models from the previous section (e.g. [44,67]). One of the models we discuss in this section was developed to describe how agents can support each other's understanding through communicative actions [44], while another offers a long-term perspective on trust development and calibration in human–robot teams [91]. Models have been developed to aid in interaction design for HRI, for instance design for interdependence [52], to establish a theoretical design framework for how products evoke social behaviour [36] and to identify how different factors influence the interaction experience [97]. We distinguish different ways of describing interaction between humans and robots in the literature proposing models of HRI, namely control relationships and social interaction including collaboration. Goodrich and Schultz identify a spectrum of different levels of autonomy of the robot relative to the human in HRI, from direct control (e.g. teleoperation) to dynamic autonomy (e.g. peer-to-peer collaboration) [39]. The models in Sect. 4.1 lie on the direct control end of this spectrum, while the models in Sect. 4.2.2 lie on the opposite end. The

[3] It should be noted that this was not the aim of Shannon and Weaver. They were mostly concerned with the technical problem of transmitting symbols accurately across a communication system, which Weaver notes has effects on other aspects of communication, such as semantics and the effectiveness with which meaning is conveyed.

models in Sect. 4.2.1 can be placed across the spectrum. In Sect. 4.3, we critically discuss the models and their shortcomings. As our aim is to focus on concepts of asymmetry and 'otherness' in human–robot interactions, the models that describe social interactions in Sect. 4.2 are most relevant to the purpose of this paper.

4.1 HRI as Control

In the models in this section, the relationship between humans and robots is one of control. The human interacting with the robot determines the robot's actions. Some models in this category aim to model teleoperation, but others see general human–robot interactions as control relations.

An example of the control paradigm can be found in Yanco and Drury [96], who propose a taxonomy intended for general human–robot interactions, from situations such as controlling an unmanned aerial vehicle (UAV) to social robots. The closest thing to a model in their taxonomy are their illustrations of various configurations of a human–robot team [96, p. 2843]. The interaction is understood as one involving control of one or more robots by one or more human operators. The human operators have different levels of interaction and shared agreement between them, while the robots prioritize, deconflict and coordinate tasks issued by the human controllers. They write: *"In human–robot collaborative systems, communication mode is analogous to the type or means of control from the human(s) to the robot(s) and the type of sensor data transmitted (or available to be transmitted) from the robot(s) to the human(s)"* [96, p.2841]. Sheridan [35] proposed modes of teleoperation control: direct control, supervisory control, and full automatic control. The direct control model allows the human to steer the robot and presents the human with the robot's sensor input through a UI. Supervisory control allows the human to formulate subtasks and monitor execution. The fully autonomous control model allows the human to formulate high-level goals. Mirnig et al. [67] propose a communication structure for human–robot itinerary requests in public places based on Shannon and Weaver's model, which we classify as control because the system's purpose is to respond to itinerary requests, which represents a device-like control relationship.

4.2 HRI as Social Interaction

In this section, we discuss both models of human–robot 'ecologies' and models of human–robot collaboration. We see collaboration as a specific form of social interaction, targeted at achieving a joint goal. While the models in the previous section represented an interaction similar to device operation, the inclusion of factors such as social context, joint goals, and autonomous and anticipatory action means that we need to consider aspects of experienced agency or alterity and thus asymmetry in the interaction. We discuss different conceptions of interaction in more detail in Sect. 5.

Social interaction encompasses social, emotional, and cognitive elements [39]. Dautenhahn [23] distinguished five different ways in which robots can be defined as social, namely socially evocative robots, socially situated robots, sociable robots, socially intelligent robots and socially interactive robots. Socially evocative robots rely on human tendencies to anthropomorphize, which is in line with the media equation proposed by Reeves and Nass [73]. Socially interactive robots, on the other hand, have high-level capabilities that enable them to collaborate with humans [4].

4.2.1 HRI as Interaction in a Social Context

Several models of human–robot communication and interaction have been developed with the aim of supporting interaction design. Several frameworks have been developed in which the robot can perform multiple roles; the robot may function both as a social agent and as a tool. The rationale for these models is that the social context is of high relevance to interactions between humans and robots. The social context, in these models, refers to relationships between the robot and other agents and the activities they undertake in a social environment. In the Domestic Robot Ecology framework [83], for instance, the social context can be taken to refer to a domestic setting in which other agents are family members and pets.

Young et al. describe the concept of *holistic interaction experience* as a way to analyse and design interaction between humans and robots and introduce the perspectives *visceral factors*, *social mechanics*, and *social structures* [97]. Forlizzi introduced the Product Ecology Framework, which proposes to study the social and physical context in which products such as robots are used [36]. Sung et al.'s Domestic Robot Ecology (DRE) framework is used to describe relationships with the environment engendered by the robot. The main factors are space, domestic tasks, and social actors [83].

4.2.2 HRI as Collaboration

Moving beyond control, some authors also characterize human–robot interactions in terms of collaboration (cf. mixed initiative interaction and dynamic autonomy in [39]). The focus is on human–robot systems that are geared towards

[4] At an even more basic level than considering robots as socially evocative, we can also consider robots as artefacts that are constructed within a human culture. As technological artefacts, they are the result of social processes; they can be described as social facts. In addition, they can also support communication between people and in this sense function as social media [37]. This means that robots that follow a control paradigm can also encode human biases and stereotypes.

achieving joint goals. The interaction advances based on task progress and the actions of the agents involved.

The collaborative control model for robot teleoperation as proposed by Fong et al. allows for human intervention in the robot's cognitive and perceptual processes. The robot can ask for the human's assistance, enabling humans and robots to collaborate as partners [35] (see also [80, p. 761]. Johnson et al. describe the Coactive System Model, which supports designing for interdependence in human–robot collaboration [52]. Hellström and Bensch propose a model of interaction that describes how humans and robots support each other's understanding through communicative actions. This can be seen as a requirement for enabling collaboration. Communicative actions by the robot seek to decrease the mismatch between what the human thinks the robot's state is and the robot's actual state. The model by Hellström and Bensch encompasses a double Shannon loop, with the addition of an extra transmission chain to loop information back to the sender. They add a factor for the general interaction context and note that the robot's inferences regarding the human's state-of-mind are influenced both by the human's communicative actions and the robot's state, in order to address the criticism that Shannon's model disregards context [44]. Malik and Bilberg propose a model for Human–Robot Collaboration (HRC) in the manufacturing domain, comprising the dimensions team composition, interaction levels, and safety implications [66]. De Visser et al. present the Human–Robot Team (HRT) Trust Model for the development of trust in teams comprising both humans and robots. The model describes how two actors engage in a process of (social) trust calibration. They propose that 'relationship equity' is an important aspect of trust building. This factor results from the multiple positive and negative experiences that an actor encounters over the course of the interaction history. During interaction, the actors adapt their trust stance (or attitude) towards the other agent. This trust stance helps the actors decide whether it is a good idea to collaborate on a certain task, and how to do so: is implicit agreement enough, or are formal work arrangements necessary? The trust stance is determined by the perceptions an actor has of another actor: these perceptions inform a risk assessment procedure (passive trust calibration). The process of active trust calibration involves the formation of a theory of mind on the part of each actor. This enables each actor to reason about the mental model of the other actor [91].

4.3 Critical Discussion of Existing Models in HRI

In this section, we identify shortcomings in current models, which we aim to address by means of a new model in Sect. 6.

The way the interaction itself is depicted differs across the models described in Sects. 4.1 and 4.2. For models that describe control relationships, the interaction is often depicted with nothing more than an arrow between the human(s) and robot(s) [35,96]. However, such a description or depiction is not informative with regards to which factors are relevant to the interaction and especially how the human is influenced by the exchange. The social and collaborative models are more detailed in this regard. The most extensive interaction component can be found in the HRT Trust Model [91]. This interaction model includes perception and collaboration components, with collaboration encompassing the subcomponents of formal work agreements, costly and beneficial relationship acts, relationship equity, and informal collaboration [91].

Something that is quite common in models of human–robot interaction or communication is that the human and robot are both depicted as agents that function in similar ways. This is especially the case in models in which we framed the human–robot interaction as collaboration. We find models in which the human and the robot are depicted as equal entities in [44,91], for instance. In their model, De Visser et al. assume advanced human-like social capabilities that robots may possess in the future. These supposed capabilities include possessing representations and an understanding of the behaviour of itself and other team members, and collaboration. At the moment, the authors note, the relationship is asymmetric and involves compensation on the human's part, as robots are unable to perform at this level [91]. In line with our discussion in Sect. 2.3, we want to stress this asymmetry. If interaction between humans and robots is depicted as two boxes with a double-headed arrow between them, this suggests that the agents are two individual entities with equal status in terms of agency. We argue that collaborative models seem most useful for modelling asymmetric agency, as they provide more detail regarding the interaction itself as well as cognitive factors and requirements.

Most models that frame HRI in terms of control or collaboration as discussed in Sects. 4.1 and 4.2.2 focus on the human and the robot in the interaction. While many of these models consider the human and the robot in isolation, the models intended for interaction design take contextual and social factors into greater consideration. For instance, the Domestic Robot Ecology framework describes how the robot invites relationships with its environment [83]. A risk of viewing human–robot interactions as something strictly involving one human and one robot is that external influences and power relationships are disregarded. In practice, the interaction is influenced by outside forces, such as functionalities provided by companies. This is the case for all proprietary aspects of platforms included in the robot's hardware and software (e.g. operating system, sensors, data processing). Intelligent virtual assistants such as Amazon Alexa typically store and process text and voice commands in the cloud, and interface with other applications [11]. Functionality that depends on external processing can play an important role in the interaction between humans and robots, which becomes especially apparent in case of failure. For instance, when the company Jibo Inc. went out of business, its servers were shut down, thereby severely limiting social robot Jibo's functionality [89]. In such cases, the robot does not function by itself, but interfaces with external entities. External influences should be made explicit in order to understand the ecosystem associated with the robot. This becomes especially important over longer periods of time and when personal privacy is impacted.

5 How Interaction and Communication are Conceptualized in HRI Models

In this section, we critically discuss the main ways in which interaction is conceptualized in the models in Sect. 4, with a particular focus on communicative aspects of interaction. We distinguish different ways of conceptualizing interaction and communication, namely in terms of sending signals (Sect. 5.1), as actions in which agents implicitly construct ideas regarding the beliefs of their interaction partner (Sect. 5.2), interaction as joint action (Sect. 5.3), and interaction as a dynamic system (Sect. 5.4). The types of interaction discussed in this section are listed in increasing order of complexity. While discussing communication as the sending of signals can be conceived of as seeing communication in terms of discrete events, interaction in Sect. 5.2 is viewed as a chain of individual communicative actions in which agents build a conceptual model of the (mental) state of their interaction partner. These two views also bring to mind a turn-taking view of communication. Viewing interaction and communication as joint action or as a dynamic system, on the other hand, implies a more continuous view of communication, in which interaction partners can monitor each other and their environment and coordinate their actions. A joint action view of communication builds on concepts such as (or similar to) Theory of Mind and offers conceptual tools to describe an embodied coordination process between agents. The dynamic systems perspective goes one step further and integrates concepts from the first three views into a view of communication as a multimodal coordination process that is described as a self-organizing system.

Although all levels of discussion are relevant, the third view is at present most useful to describe situations in which robots are implemented to achieve shared goals in human–robot teams, which is why we focus on this view in the model we propose in Sect. 6. It is at present not easy to combine cognitive-level reasoning processes regarding common ground, for instance, with the dynamical systems approach, as noted by Dale et al. [22, p.62]. We argue that the joint action approach provides relevant conceptual tools to describe coordination, which is why this approach is the focus of the present paper. Joint action implies that communication is a participatory process in which participants have shared goals. The joint action perspective is especially useful for HRI, as it offers concepts to think about the way collaboration can be achieved.

5.1 Interaction and Communication as Sending of Signals

Communication and interaction can be discussed in terms of the exchange of signals and cues. Such a discussion falls within the semiotic tradition of communication theory. Peirce and Saussure are generally recognized for their contributions to the study of signs. In Saussure's semiotics, a sign consists of the signifier and the signified. These concepts cannot be disentangled [94]. The signifier is the 'sound-image', and signified is the concept. A sound-image is the combination of what one hears and sees in response to a spoken word. Signification is the process of making use of signs with their associated meanings [62]. In Peircean semiotics, the symbol is treated as a process (semiosis) with three components, namely the sign (representamen), object and interpretant. The *sign* is the form of the symbol. What is represented by the sign is called the *object*. The *interpretant* is an effect on the person that forms the relation between sign and object [84]. Saussure noted that while the meaning of signs relies on convention, signs can also be interpreted in different ways. Signs are used in an intentional way, and for a sign to be a sign it has to be interpreted as such. Peirce, in contrast, saw signs as a means for people to think, communicate and make their environment meaningful. This does not require the sign to be part of intentional communication [94].

Typologies and taxonomies have been proposed to classify signs and cues for HRI and conversational agents. For instance, Hegel et al. [43] propose a typology of signals and cues in HRI, and distinguish between human-like and artificial signals and cues. Signals are designed to provide information, while cues (such as motor sounds) are all those features that can potentially be informative but were not necessarily explicitly designed as such. We note that the distinction between natural and artificial can cause confusion when applied to robotics. All robot signs are (at least currently) artificial. There are degrees of human-likeness (and animal-likeness, plant-likeness). Human-likeness is a spectrum, not merely a property. Feine et al. [32] propose a taxonomy of social cues for conversational agents (CA) based on a systematic literature review. This proposed taxonomy contains the main categories verbal, visual, auditory and invisible cues. The authors define a cue as *"any design feature of a CA salient to the user that presents a source of information"* [32, p.141] and a social signal as *"the conscious or subconscious interpretation of cues in the form of attributions of mental state or attitudes towards the CA"* [32, p.141]. These definitions are based on a view of cues as all those features that can provide information to the user. Cues only become social signals if the cue leads to attribution of sociality to the CA by the user. A social cue, then, is a cue that actually provokes a social reaction on the part of the user, where a social reaction is a reaction to a conversational agent that would also be appropriate if it were aimed towards another human [32]. Such a view of cues and signals is also useful for robotics, and avoids the question of whether we should regard the robot as a social agent. This view focuses the discussion on the perception of the signal as a social signal by the human.

Communication, at the level of analysing signals and cues, is used to affect the behaviour of an interaction partner or to convey information. The signal or cue is discussed as a discrete event that carries information. For instance, Hegel et al. [43] provide a table in which they note that an artificial signal such as an LED can convey activity information, for instance, while a humanlike cue such as body size conveys dominance information. While cues and signals convey information, we would like to note that context is also relevant and, as also noted by Feine et al., signals and cues do not occur in isolation [32]. Analysis at the level of signals and cues can serve as a basis for discussion, but needs to go further.

5.2 Interaction and Communication as Communicative Action

Bensch et al. define interaction events as tuples of *perceived information* and *associated actions*. Interaction events link together in chains to form interaction acts [4]. Hellström and Bensch define the term *communicative action* as *"(…) an action performed by an agent, with the intention of increasing another agent's knowledge of the first agent's SoM"* [44, p. 115], with SoM referring to the agent's state of mind. The advantage of Hellström and Bensch's definition is that it can be used to "computationalize" communication; it lends itself to being written in algorithmic form, in which the agent is able to compare the contents of its own belief system to the one it estimates a second agent to have regarding the first agent. While this can be advantageous for AI and robotics applications, we note that this could also be a risk if important factors in the interaction are not (or incorrectly) captured and processed. Hellström and Bensch's definition also applies to the model they propose [44].

The concept of Theory of Mind (ToM) is central in such a view of communication processes. Krämer et al. write that *"Theory of mind (ToM) is the ability to see other entities as intentional agents, whose behavior is influenced by states, beliefs, desires etc. and the knowledge that other humans wish, feel, know or believe something"* [58, p. 54]. Krämer et al. argue that the concepts of common ground, ToM and perspective taking are similar, as these concepts propose that humans have implicit knowledge regarding how other minds work, which they use as a basis for mutual comprehension, and that they enhance mutual knowledge through grounding processes [58]. In the context of robotics, De Visser et al. define Theory of Mind as *"An actor's (e.g. actor A) estimation of another actor's (e.g. actor B) mental model of that actor (e.g. actor A)"* [91, p. 461]. There can be different levels of ToM, in which each level encapsulates the former: for example, if Level 1 contains A's individually held beliefs, then Level 2 can contain A's estimation of B's beliefs, Level 3 contains A's estimation of B's estimation of A's beliefs, Level 4 contains A's estimation of B's estimation of A's estimation of B's beliefs, and so on. Such a conception of Theory of Mind has some similarities with the concept of ToM for interpersonal communication, but is clearly a much reduced form. Robots do not attribute mental states to others in the sense that humans do, but may be equipped with algorithms and mechanisms to estimate emotions or beliefs held by people.

5.3 Interaction and Communication as Joint Action

One perspective that is discussed in the context of HRI views communication and interaction between a human and a robot as a form of joint action. This perspective treats (linguistic) communication and coordination of actions as similar processes. Joint action can be defined as *"any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment"* [90]. The term can also be found in symbolic interactionism, which is part of the sociocultural tradition of communication research. Social acts consist of the relationship between a gesture by one actor, a response by another and a result. Blumer describes joint action as an *interlinkage* of social actions [63]. The concept has been studied in cognitive psychology and has previously been applied to interaction with artificial agents [19] and robots [47]. The current section builds on the previous one, as coordination processes involve more than awareness of the mental states of others.

Clark conceptualizes communicative acts between humans as participatory acts: the communicative action by one individual who signals and another who recognizes the signal is a joint act [15]. This view expresses that the actors are both actively involved in constructing the meaning of the information exchange. Clark discusses joint activity as an activity that is performed by two or more participants. These participants have activity roles, which helps establish a division of tasks. Participants strive to achieve (joint) public goals, but may have private goals as well. Participants in the joint activity have prior mutual knowledge or common ground, which accumulates over the course of performing the joint activity [14]. Common ground refers to mutually held beliefs, see also Sect. 7.6. Joint action involves coordinating content and process, which are themselves interrelated. Clark calls joint actions a paradoxical concept, as a group of humans does not itself intend to perform actions. Instead, individuals perform participatory actions. These actions are performed when coordination is required in order to meet common goals [13]. Clark argues that joint activity cannot be separated from language use and views language use itself as a form of joint activity. People use language to coordinate their actions, and the language used does not make sense apart from the context of action in which it is applied [14].

Krämer et al. [58] write that there are several basic communication capabilities that humans are used to and that will also be required for successful human-agent and human–robot communication, notably social perspective taking, common ground, and Theory of Mind. Coordination devices to support joint action include explicit agreement, precedent, convention, and joint salience. Whether something is salient to all participants in an interaction, depends on their common ground. Participants can usually assume the solvability and sufficiency (with respect to the available information) of a current coordination problem, when one of the participants proposes the problem themselves [13]. Vesper et al. [90] discuss coordination mechanisms between humans in intentional joint action. Participants in the joint action use mental representations of the joint action goal and the task in order to monitor task progress. The co-actors share information throughout the task using mechanisms such as shared gaze, predicting the actions of the other actors involved, sensorimotor communication, and haptic coupling. They also express emotions and interpret those of others. The mechanisms they use for joint actions are coordination smoothers (e.g. synchronizing actions), communicated and perceived affordances, and cultural norms and conventions [90]. Mutlu et al. [68] describe coordination mechanisms established in cognitive science, such as joint attention, action observation, task sharing, coordination of actions, and perception of agency. Mechanisms such as gaze, action observation and conversational repair have also been investigated in the context of robotics [68].

Clodic et al. [16] present a framework for interaction between humans and autonomous robots with the aim of achieving human–robot joint action. The authors align Pacherie's theory of joint action with a three-layered robot architecture. Clodic et al. write that joint action not only requires individual agents to have common goals and be able to execute plans and actions, but that they must also be able to coordinate their individual plans. This coordination of subplans needs to occur prior to as well as during execution. This requires the capacity to monitor and predict the actions (and intentions) of one's partner. They note that motivational uncertainty (do the goals of the other agent align with mine?), instrumental uncertainty (how do we achieve the goal?) and common ground uncertainty (are we both on the same page regarding the goal and actions to take to get there?) can negatively affect mutual predictability. The authors note that self-other distinction would be a required capability for the robot: it would have to maintain 'mental' models of both itself and of the human [16]. Compare this to the concept of Theory of Mind as described in Sect. 5.2.

5.4 Interaction and Communication as a Dynamic System

Interaction can also be conceived of in a different way, namely as a dynamic system in which the interaction partners are simultaneously monitoring each other and their environment for cues and signals. Dale et al. [22] propose dynamical systems theory as a framework for a more comprehensive theory of human interaction. The authors write that many theories and concepts have been proposed to describe aspects of human interaction, such as perspective taking, joint action, ToM, and mimicry. In order to understand how these different accounts and types of processes form a multimodal coordination process, they propose that human interaction functions as a synergetic, self-organizing system. Sandry [76] refers to human-animal communication to argue that communication and interaction are more like a dynamic system than a dialogue with strict turn-taking. She writes: *"Communication operates as a dynamic system during this type of embodied communicative situation, and signals between communicators overlap as human and [animal] continually reassess each other's position, perceived intention and likely subsequent action"* [76, p. 40]. In this dynamic system, sometimes the meaning of an individual communicative act can be understood easily, while in other cases the meaning can only be derived from other communicative acts in context. This can also be applied to robotics: consider the situation of monitoring a robotic system in a manufacturing context. The interaction does not follow a script, but rather consists of the human paying attention to the system, with events (such as being alerted by the robot if something goes wrong) provoking action on the human side.

The idea of interaction as a dynamic system is echoed in the concept of interaction fluency. Hoffman [46] proposed fluency metrics to assess how fluently a human–robot team interacts. By making an interaction more fluent, it is proposed, one moves away from a strict turn-taking interaction towards an interaction in which the robot starts to anticipate on human actions, allowing for overlap between human and robot actions instead of only one agent acting at a time. Conceiving of interaction as such a dynamic system is one step beyond mixed-initiative interaction, as it not only involves considering who takes initiative: it also requires anticipating the other agent's actions.

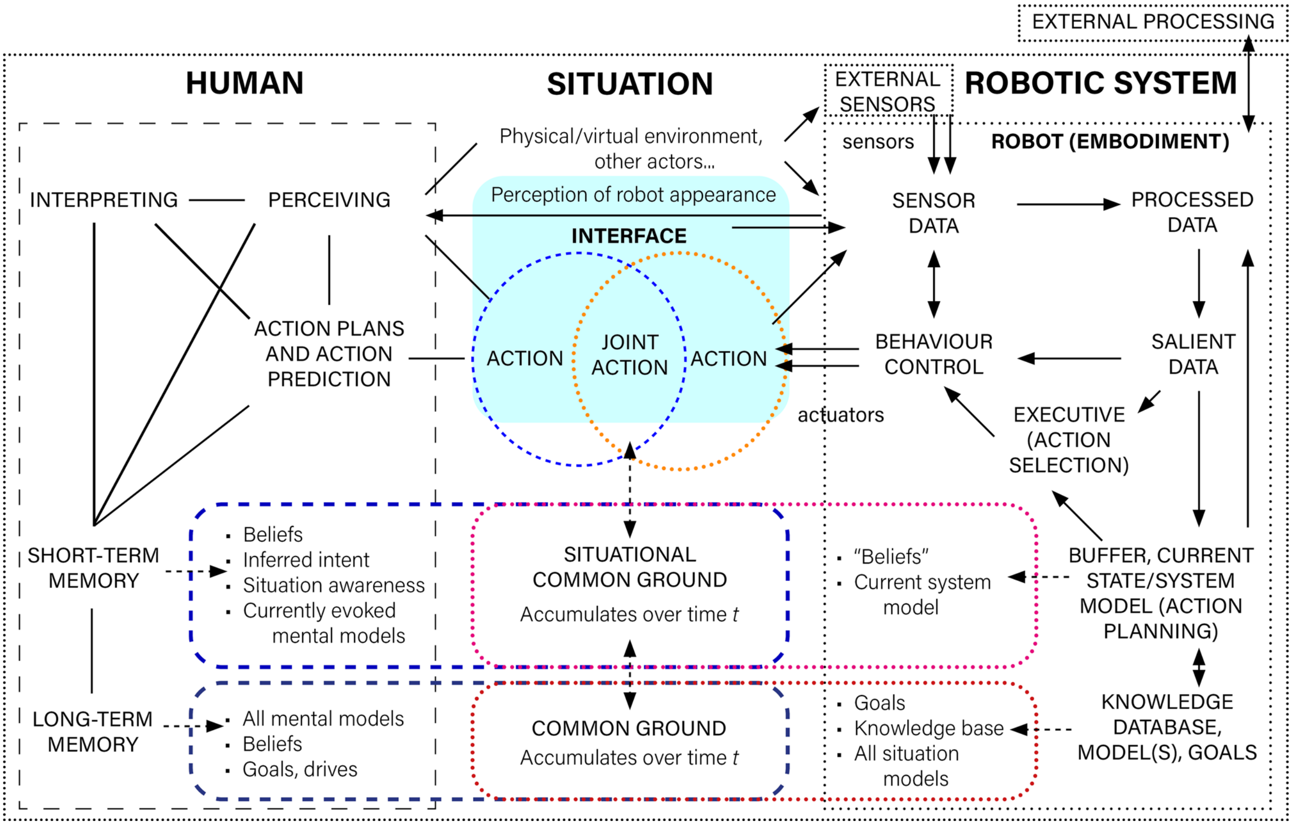

Fig. 1 Here we propose an Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI). This is a model with global similarity between the components/organization on the human side and the robot side, which allows for comparison. However, the processes that occur are different, to allow for identification of differences between the two agents. The 'interface' includes everything that allows for communication/information exchange, including observed actions and observations of embodiment. **On the human side:** Interpretation of percepts leads to new beliefs, which are added to short-term memory. When actions are executed, their effect is predicted and the action execution is monitored. **On the robot side:** Moving from data to processed data requires feature extraction and data processing. Processed data is then further processed; algorithms determine which information is salient through attention mechanisms, data fusion, and matching processed data to patterns in the database. Between the buffer and the knowledge base, data updating and retrieval processes occur. Solid-line arrows indicate processes, dotted arrows indicate a theoretical connection

6 An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI)

In this section, we propose an Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI). Humans and robots are very different entities. To model interaction or communication between these different types of entities, they should be depicted in an asymmetric way. Robots can also be subject to outside control (external influence). A model of human–robot interactions in which the differences in functioning between human and robot agents are highlighted, can help identify possible mismatches between robot capabilities and human expectations. Based on the discussion of the models of communication and interaction between humans and robots found in the literature, we propose that a model for collaboration between humans and robots that depicts the human and the robot as different types of agents (thus, an asymmetric model) will be useful for identifying how the human and the robot operate differently, and predicting mismatches between robot capabilities and human expectations of those capabilities. A joint action perspective on interaction emphasizes the collaborative co-construction of meaning by the agents involved, and is more suited to a scenario of human–robot collaboration in which the robot supports human work, which is why we chose to use this perspective on interaction from Sect. 5.

The model as depicted in Fig. 1 shows a robot architecture with common components and depicts the processes by which a human and a robot may establish common ground. This model will be used as a basis to compare the human and the robot sides of the model in Sect. 7. Moreover, based on the identified differences, we propose design recommendations in connection with a version of the model, as illustrated in

Fig. 3. The models and design recommendations are mainly intended for interaction designers in the field of HRI.

6.1 The Model

We propose a model based on Kincaid's diagram of the components of the convergence model of communication [54], Christensen and Hager's diagram of the robot sensing process [80, Ch.4] and Bauer et al.'s diagram of mechanisms for robot joint action [3], see Fig. 1. We draw a parallel between theory regarding joint action and Kincaid's convergence model of communication. Kincaid views communication as a process that has the aim of arriving at mutual understanding, which is achieved through creating and sharing information. His model contains components such as mutual agreement and collective action. The terms used by Kincaid are similar to the vocabulary in the literature on common ground [5]. Here, we use the term *joint action* in line with the HRI literature. Kincaid's concept of mutual understanding can be related to the model and definitions by Hellström and Bensch, as well as the concept of ToM as used in the robotics and AI literature (see Sects. 5.2 and 5.3).

The proposed model in Fig. 1 includes indications of processes that could occur in communication in human–robot teams. However, we note that the model of the human side is highly abstracted and incomplete. We do not propose that this is an accurate, complete depiction of human processing, but we still include a sketch of the processes on the human side to allow for comparison between the human and the robotic agent. The model in Fig. 1 is structured in such a way that both agents are as similar as possible on a high level, while still indicating differences in terms of processes. The model visually expresses the asymmetry between humans and robots. The robot side displays processes that may be lacking in many robots. The point is that the model depicts the human and the robot with respect to similar processes, allowing for comparison, while still being expressive of major differences between the agents that are relevant for interaction design in HRI. (If the model were completely asymmetric, with no correspondence of components or processes on the human side with components on the robot side, the two sides could not be compared. The asymmetry arises partly by means of comparing the two sides.)

We now discuss the components listed in the 'human' component of the model, followed by a discussion of the situation component and the processes occurring on the robot side.

6.2 Discussion of the Proposed Model: Human

The processes on the human side of the model are mainly intended for illustration, drawing parallels, and illustrating contrasts with processes occurring on the robot side. The processes on the human side are based on the model of human–human communication by Kincaid and Endsley's models of situation awareness, which models human-technology interaction. Endsley proposes a model of situation awareness (SA) and discusses its role in the context of human decision-making. SA plays a role in applications ranging from air traffic control and tactical systems to decision-making in everyday activities. Endsley distinguishes three levels of situation awareness: (1) perceiving information, as well as (2) comprehension/understanding, and (3) forecasting future states to aid decision-making. Endsley proposes a model regarding the interrelation between a person's goals, mental models, and situation awareness, as well as a model depicting the mechanisms that are important in establishing situation awareness [30]. The components long-term memory, short-term memory, and the terms situation awareness and mental models are derived from her models. The long-term memory component expresses that the human does not have all their knowledge and experience at hand to apply to a new situation, but instead retrieves a subset of knowledge in response to the current situation.

As mentioned earlier, Pickering and Garrod [72] proposed that human comprehension and production processes should be understood as tightly interwoven, which supports human agents in predicting the actions of their interaction partner as well as their own actions, interweaving actions within a coordination process, and achieving joint action. Separating production and comprehension would lead to a model that Pickering and Garrod [72] refer to as a "cognitive sandwich" (as coined by Hurley [49]) in which cognition is sandwiched between perception and action. While this "cognitive sandwich" may well exist in many robot architectures, we assume in the model that perception, action and interpretation are tightly coupled on the human side, as denoted by lines (rather than the arrows indicating one-way processing on the robot

[5] The combination of the terms *mutual understanding* and *mutual agreement* in Kincaid are similar to the term common ground, while Kincaid's term *collective action* is similar to the term joint action. Kincaid defines mutual understanding as *"the combination of each individual's estimate of the other's meaning which overlaps with the other's actual meaning. In other words, mutual understanding is a combination of the accuracy of each individual's estimate of the other's actual meaning"* [54, p. 32]. Kincaid describes mutual agreement the following way: *"When two or more individuals believe that the same statements are valid, they become true by consensus, or mutual agreement with some degree of mutual understanding"* [54, p. 31] while Clark describes common ground as *"…The sum of (…) mutual, common, or joint knowledge, beliefs, and suppositions"* [12, p. 93]. Both the definitions by Kincaid and the one by Clark refer to mutually held beliefs. We refer to this concept as *common ground*. Collective action (Kincaid) is the *"(…) result of the activities of two or more individuals (A and B), built upon a foundation of mutual agreement and understanding"* [54, p. 31], while joint action (Clark) is activity performed by two or more individuals based on common ground and joint action goals.

side). In contrast, the processes on the human side are not one-way but can mutually influence each other.

When human behaviour or human cognitive processes are modelled, these models are necessarily limited; the model developer is required to make assumptions regarding human behaviour, perception and cognition. While models can be useful, as in the context of this paper, it remains important not to lose sight of people's embodied, individual and varied experience. It should be noted that every human is different, has individual needs and abilities as well as a personal background and identity. This variety and these differences need to be kept in mind rather than assuming an imaginary 'general human'. Spiel argues that we should *"appreciat[e] the plurality of human bodies instead of assuming a specific embodiment"* [82, p. 2] in the context of technology design for embodied interaction. The same goes for human cognitive processing, action and perception capabilities, and other components of the model in Fig. 1. The components listed on the human side are a limited selection and real-world human capabilities go beyond what is included in the model, as they are far more complex and more varied than what can be captured in an illustration.

6.3 Discussion of the Proposed Model: Robot

The model on the robot side is based on a model of the robot sensing process [80, Ch.4], a model of mechanisms enabling joint action for robots [3] and theory on cognitive architectures. For a discussion of how cognitive architectures can be structured, we refer to the survey by Kotseruba and Tsotsos [55] and the review by Chong et al., who compare the cognitive architectures SOAR, ACT-R, ICARUS, Beliefs-Desire-Intention (BDI), the subsumption architecture and CLARION [10]. The goal of research on cognitive architectures is to model the human mind and achieve human-like intelligence. Cognitive architectures are also developed and implemented for robotic systems, see for instance [87].

With regards to the model proposed in Fig. 1, cognitive architectures and computational models of human(like) behaviour are useful both for developing the system architecture on the robot side of the model as well as for maintaining representations of the state of human interaction partners in the robot's memory, in order to facilitate collaboration. Perception, attention, action selection, memory, learning, reasoning, and meta-cognition are the main features of cognitive systems (and have been developed further than, for instance, emotion and creativity) [55], and these are the features we focus on in the proposed model. Based on the research on cognitive architectures, we included short-term and long-term memory components (see, for instance, the SOAR architecture in [10]). In the next sections, we provide more details regarding robot sensing, knowledge representations, reasoning, learning, action selection and actuation. The discussion will focus on the robot's awareness of humans and their behaviour, as this is relevant in a communication process.

6.3.1 Sensing

In the model in Fig. 1, the sensing process consists of the components *Sensors* and *External Sensors* that collect sensor data, the component *Processed Data,* and the component *Salient Data*. The sensing and perception process in robotics has been described as a process that involves updating an existing partial world model with sensor data. This process involves feature extraction, matching or associating data with an already existing model, updating and integrating new knowledge in the model, and prediction of future states (which influences data matching) [80, p.88]. Yan et al. [95] identify feature extraction, dimensionality reduction and semantic understanding as key components to social robot perception systems.

According to Christensen and Hager [80, Ch. 4], sensors for robotics applications can be classified in the following way: tactile, haptic, motor/axis, heading, beacon based, ranging, speed/motion, and identification. The types of sensors installed on a robotic system depend on the application; for instance, medical robots and industrial robots will need different sensors than assistive robots intended for social interaction. We can make a distinction between sensors for interoception and exteroception. Interoception refers to sensing the robot's state (e.g. motor currents). Exteroception refers to sensing the external world (e.g. distance to an object) [80, Ch. 4]. Sensors can also be classified as passive (does not emit energy) or active (emits energy in order to sense) [80, p.452]. They can also be on-board or external (consider, for instance, the concept of the Internet of Things; in such a scenario, the robot may have access to external sensors and devices).

The main types of signals that social robots make use of for interaction with humans are based on visual, audio, and tactile interaction modalities, as well as ranging sensors [95]. Visual-based signals can be captured using 2D and 3D cameras (using depth information with RGB-D cameras or stereo vision). Audio data can be captured using microphones for subsequent speech recognition, another key aspect in human–robot interactions [86]. Data from different sensors and interaction modalities (vision, audio, touch) are combined and subjected to processing (for instance, by means of computer vision methods using Hidden Markov Models (HMMs)) [86]. In order to make sense of high-dimensional data, statistical techniques (such as principal component analysis) can be used for processing the data and extracting features from the data in a lower-dimensional feature space, after which the data can be used for applications such as object recognition [95]. Recognizing humans and features of humans and their behaviours are of high importance for social interaction with robots. Research on computational HRI has focused on topics such as detection of body pose, face recognition, activity and gesture recognition, and interaction engagement [86].

The *Salient Data* component is included in the model to indicate that processed data may be further processed to identify the features in the environment that are most relevant to the interaction. Unimodal feature extraction is often not robust enough; therefore, multimodal feature extraction methods can be used, in which data from separate modalities are combined into a saliency map [95] or measures of object saliency (e.g. [87]).

6.3.2 Knowledge Representations, Reasoning and Learning

A suitable knowledge representation is required for reasoning about information and storing it. The field of knowledge representation is concerned with finding representations that are adequate in an epistemological sense (represent referents in the environment in a compact, precise way) and in a computational sense (that is, efficient) [80, Ch. 9]. The formalisms used for knowledge representations and making inferences are mainly based on logic and probability theory [80, Ch. 9]. Reasoning has specific issues in robotics applications as compared to other types of knowledge-based systems. Robots are embedded in dynamic environments and have to interpret and respond to environmental information (partially) autonomously in near real-time. Approaches that try to remedy these issues include fuzzy logic approaches and embedding time constraints within the robot's architectural design [80, Ch.9] (see also the KnowRob system for an example [85]). Learning on the robot side can occur in different ways. Kotseruba and Tsotsos describe learning as *"the capability of a system to improve its performance over time"* [55, p.50], based on experience. They distinguish between declarative and non-declarative learning, where non-declarative learning encompasses the learning mechanisms perceptual, procedural, associative and non-associative learning [55]. One specific type of AI is machine learning. Hertzberg and Chatila define machine learning in the context of robotics as *"the ability to improve the system's own performance or knowledge based on its experience"* [80, p. 219]. Methods include inductive logic programming, statistical learning, and reinforcement learning. Learning can be supervised or unsupervised [80, Ch. 9].

A robot architecture can also be designed to support some level of metacognition. Metacognition includes introspective monitoring of the robot's status and processing (e.g. self-observation) and Theory of Mind (ToM) [55]. In order to accommodate social interaction with humans, robots can be equipped with mechanisms based on ToM (which means, in the context of cognitive architectures, that the system infers others' mental states and uses this information for decision-making [55]) and ways to explicitly model humans and human behaviour. Cognitive architectures have been proposed that draw on the concept of ToM, in order to infer human intentions from goal-directed action [86]. However, most social robots are far from full ToM. At present, research has been conducted on the development of capabilities such as parsing human attention, which may aid in the achievement of human–robot joint attention, and predicting human action in order to be able to anticipate on it [86]. Hiatt et al. [45] review different ways of modelling human behaviour that can be implemented in a robotic system with the aim of enabling the robot to understand a human teammate's behaviour. They write that computational approaches (such as conventional machine learning approaches) can be useful in situations in which rational, 'ideal' or 'typical' performance by humans can be assumed, but this leaves little room for human error or deviation from set norms, although such deviations are to be expected in human–robot collaboration. They also discuss computational/algorithmic approaches such as HMMs and the cognitive architecture ACT-R/E [45].

6.3.3 Action Selection and Actuation

Action selection can occur dynamically (choosing one option from a set of alternatives) or in the form of action planning (as is common in traditional AI) [55]. Planning problems are usually described as sets of states with actions that can induce transitions between states. The goal is to find a suitable series of actions from the start to the goal state. Action planning can involve working towards a common goal for efficient human–robot collaboration [3]. Robot planning uses planning methods that make use of formalisms from logic and probability theory to complement motion planning [80, p. 219]. In the research area of computational HRI, fluent meshing of actions, human-aware motion planning, object handovers, and collaborative manipulation are important research foci for robot action planning [86].

Motion trajectories by the robot should be possible to execute; therefore, motion planning needs to take the robot's kinematic constraints into account. Aside from achieving task goals, robot actions such as robot motion can communicate intent to an interaction partner or observer, whether or not the action is planned to be communicative. Motion can also have a communicative aspect: instead of purely functional motion planning, generating motion that is legible and/or predictable to human interaction partners can also be considered [25]. Social robot navigation has a social component as well, as demonstrated by the research topics of approaching humans, navigating alongside people, and human-aware robot navigation [57,74,86]. The use of gestures and gaze cues, proxemics, haptics, affect, emotions, and facial expressions have been studied as nonverbal behaviour that can be implemented in robots for communication [86]. The robot can also use other interaction modalities as part of a communication process, for instance by making use of auditory signals (see also Sect. 5.1) or changing the state of a graphical user interface that is part of the system.

6.4 Discussion of the Proposed Model: Situation and Interaction

In the model, the *Situation* refers to the current proximate physical and social environment (the interaction context), the current constellation of agents, objects and environment, close to each other in space and time. It includes other agents or actors that may be involved or referenced in the communication process.

Joint action consists of actions involving both agents that have the aim of establishing common ground or achieving shared goals. Joint action is a subset of all actions, including those actions that advance the human and the robot in their joint action goal. *Situational common ground* is the subset of the interaction partners' beliefs and goals that are shared in the current situation. We included the component situational common ground as something separate from common ground, based on Endsley's theory on situation awareness, which holds that not all information is in consciousness; this is true only for a subset of information and mental models.

Joint actions are a subset of all actions carried out by the participants in the interaction. These actions are the components of larger joint activities (cf. Clark, [14]) and move the participants closer to a desired goal state. Clark differentiates between a joint act and a joint action. The former is discontinuous, while the latter is a continuous coordination process. Clark distinguishes phases as the distinctive elements that make up joint actions and that allow them to be coordinated, defining phases as *"a stretch of joint action with a unified function and identifiable entry and exit times"* [13, p.83]. Examples of joint actions are giving a person a handshake or asking someone a question. In the proposed model, no distinction is made between actions and communicative actions. However, a detailed look at research in the semiotic tradition and the work that has been done in HRI on classifying signs and cues can be useful to specify the communicative aspects of actions further.

Clodic et al. identify three levels of uncertainty, namely instrumental uncertainty (related to joint action), common ground uncertainty (related to common ground) and motivational uncertainty [16]. We can identify these levels of uncertainty in the model. Instrumental uncertainty occurs on the levels of action and situated common ground. Common ground uncertainty and motivational uncertainty both occur on the levels of situational common ground and common ground. The robot does not have 'personal' goals. This may result in increased motivational uncertainty on the human side regarding the intentions of the developer of the robot, its software, or owner, if the motivations/goals of the robot developer are not communicated.

Humans and robots can only have reduced common ground as compared to the common ground shared by humans. If the robot can only sense and act, the common ground factor in the model would become irrelevant, and instead of "joint action", we might label the aggregate of human and robot action as a "collection of actions" instead.

Participants in the interaction have internal goals or goals that have been defined externally. Participants are trying to achieve goals while engaging in joint activity, most notably the *domain goal* in Clark's words, yet participants can also have procedural goals, interpersonal goals and private agendas [14, p.34]. In human–human communication, high-level goals are usually internally defined (and then possibly negotiated), but this is not the case for robots. High-level goals may be externally dictated by human interaction partners or the company or companies that produced the robot and its components. Subgoals, on the other hand, might be either external or internal, derived from high-level goals (e.g. moving to intermediate location B while moving from A to C).

Having a 'joint intention' or a common goal refers to a joint, participatory aim that is shared across participants, *"a joint commitment to perform a collective action while in a certain shared mental state"* in the words of Cohen and Levesque [19]. The notion of joint goals, or working towards achieving a common goal, is not necessarily useful in all cases, especially if the robot is intended for social interaction and/or operating in a (semi-)public space. For instance, if a (human) visitor to a conference approaches a humanoid robot and starts waving in front of it and muttering phrases to it to see if it will respond, this behaviour could be said to have a goal on the human side (even if subconscious), namely to entertain themselves and figure out what the robot can do, but it cannot really be said to constitute collaboration or 'working together'. Joint action arises only when the action is acknowledged or responded to by another agent, and the goals of both agents align. The robot's high-level goals are defined externally, but lower-level goals (such as moving to intermediate location B while moving from A to C) can be defined internally.

One may read the agents as acting in a very goal-directed way on the basis of the preceding text (in the way of Saussure, instead of Peirce). However, a view of actions and signals as supporting a process of reflection is not excluded, and actions and communication can also be viewed in the model as a means of thinking. For instance, consider a case in which a robot pushes over a stack of blocks repeatedly and observes what happens.

6.5 Practical Example

In this section, we walk through the model using the practical example of lexicon learning. We will shortly elaborate on the example. The human teaches the robot new words by pointing to objects on a table and naming the objects. The robot stores representations of the object and the words the human uses to name those representations (accumulation of common ground). After the teaching phase, the human asks the robot to name the objects on the table that the human points at (joint action). In this fictive example, consider the robot to have a moveable head and to have pointing detection, speech recognition, basic object recognition, and face recognition functionality.

We work this out in a script form in which each robot action is specified. Only human perceptions, thoughts and actions are included. We do not presume to guess the human's inner workings, but propose one possible option for what the human may infer based on robot actions and other events. This depends also on other factors, for instance whether the human is an expert user or a novice. In practice, the expectations of (multiple different) human interaction partners can be elicited in the context of interaction experiments by means of methods such as think-aloud and post-experimental questionnaires or interviews. With regards to the interaction component, common ground uncertainty is included at relevant points. Note that we discuss one action-response pair, so a single joint action.

What can be useful about working out such a script, is that it forces the developer to be very specific regarding expected human thoughts and actions, which yields hypotheses that can be tested. It can also help with identifying whether the robot's behaviour needs to be modified.

The items labelled **Human (X)** indicate an action or process on the human side, while the items labelled **Robot (X)** indicate an action or process on the robot side. They are presented here in a sequential way, although some actions that are listed as sequential can also co-occur at the same moment. The italicized items marked with quotes are thoughts or verbalized human thoughts, depending on whether the items were elicited by means of brainstorming by the researchers (as they are in this case) or the method of think-aloud.

Human common ground uncertainty (1) Uncertainty on the human side before naming the cup

- Instrumental uncertainty: *"What can the robot do? How will the robot act?"*
- Common ground uncertainty: *"What does the robot know?"*
- Motivational uncertainty: *"Does the robot have the goal of learning the names of these objects? …I suppose so, the researcher told me?"*

Human (1) Perception

- *"I see a yellow robot with large eyes, a torso and two arms"*

Human (2) Interpretation

- *"Cute! I guess it is looking at me. I wonder what it can do."*

Human (3) Action

- The human points at a cup. The human pronounces the word *"cup"*.

Robot (1) Data is captured by the robot's sensors

- Audio is captured by the microphone.
- Image data is captured by the robot's camera.
- Depth information is captured by the robot's depth sensors.

Robot (2) Feature extraction

- A pre-trained image processing algorithm identifies that there are three objects in the camera view that do not have a stored label associated with them. A pre-trained object recognition algorithm recognizes a human face and a pointing hand.
- Speech recognition software recognizes the word *"cup"*.
- The direction in which the hand is pointing is inferred using the video and depth information, and stored as the approximate pixel area on the video image.

Robot (3) Attention and data fusion

- The object **[object1]** that the human pointed at is inferred.
- The location of the human face is inferred.
- The speech recognition result *'cup'* (semantic label) is associated with the pixels from the image that were labelled **[object1]**.

Robot (4) Action selection

- The robot looks in the direction of the object that the human is pointing at.

Human (4) Perception

- *"The robot is looking at the object"*

Robot (5) Buffer: storing data in short-term memory

- The semantic label *'cup'* and **[object1]** are placed in the buffer.
- The location of the human face is stored.

Robot (6) Matching: storing data in long-term memory

- **[object1]** and *'cup'* are stored in long-term memory.

Robot (7) Action selection based on successful storage of item in short-term memory

- After a delay of 2 seconds, the robot moves its head in the direction of the human face.

Human common ground uncertainty (2) Uncertainty on the human side after naming the cup

- Instrumental uncertainty: the robot acknowledged the human's action when it looked at the object. *"The robot looked at the cup when I pointed at it, so it must have noticed what I pointed at."*

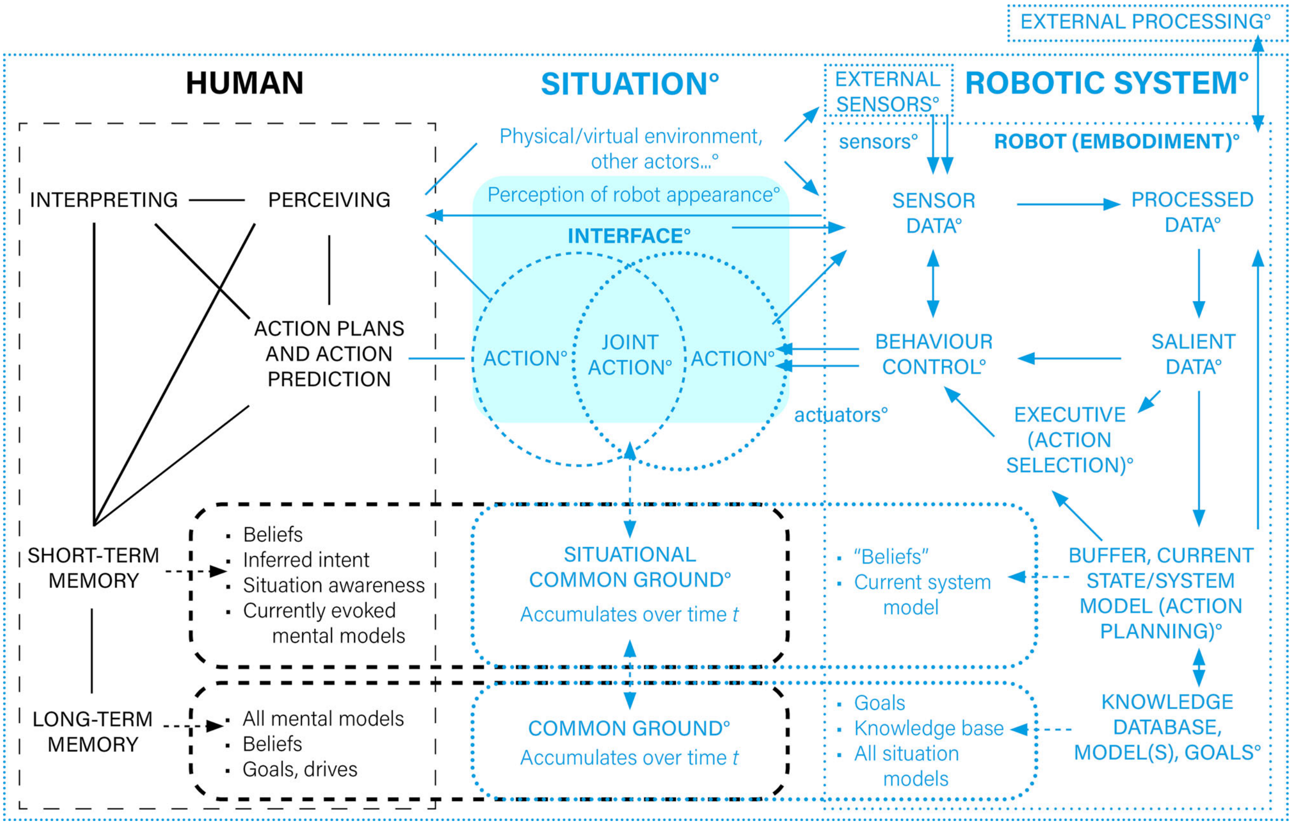

Fig. 2 Marked in blue: Possibilities to influence. It is possible to directly influence all components and processes on the robot side by reprogramming it. However, it is not possible to directly influence the human side, except on the level of input/output: it is possible to give the human different information or influence the way the human is able to execute actions. Note that it is possible to influence human processing indirectly, as things such as hormones, food, and attention all influence the way processes on the human side operate. It is also possible to change a person's belief system by changing the information environment they are exposed to

– Common ground uncertainty: *"Did the robot understand that the object is called a cup? Does the robot already know that the object is a cup?"*
– Motivational uncertainty: *"Is the robot currently trying to infer the object name?"*

After this interaction, the common ground between human and robot can be constructed as follows:

Beliefs Human (ToM Level 1) Human knows *"cup"* is associated with the object cup.
Beliefs Robot (ToM Level 1) Robot inferred that *'cup'* is associated with **[object1]**.
Beliefs Human (ToM Level 2) The human does not know if the robot knows the object is a cup.
Beliefs Robot (ToM Level 2) Robot inferred that human calls **[object1]** *'cup'*.
Beliefs Human (ToM Level 3) The human does not know if the robot knows that the human does not know if the robot understood it is a cup.
Beliefs Robot (ToM Level 3) The robot did not acknowledge that the object is a *'cup'*, so the robot may infer that the human does not know that the robot knows **[object1]** is a *'cup'*.
Situational common ground (after interaction) Both human and robot associate the object with similar labels (*"cup"/'cup'*), but the knowledge that the robot has associated the object with the label is not common ground. The robot should use this information to communicate that the word *'cup'* is now common ground, or confirm otherwise.

7 Design Recommendations

Norman identifies seven design principles for interaction design, namely discoverability, feedback, a conceptual model, affordances, signifiers, mappings and constraints [69]. These principles point to the importance of making sure a person interacting with a product or interface is able to determine which actions are currently possible, what the

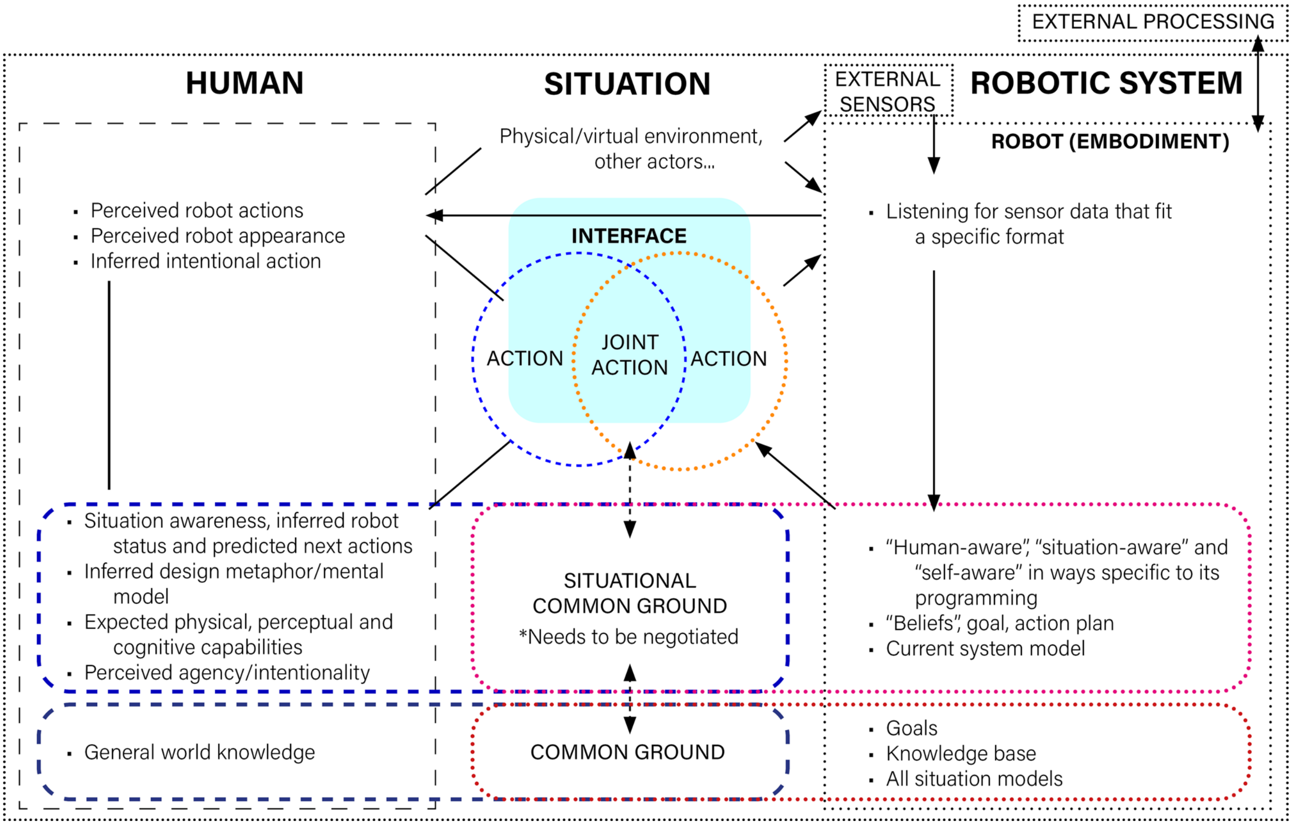
EXTERNAL PROCESSING
HUMAN
SITUATION
EXTERNAL SENSORS
ROBOTIC SYSTEM
ROBOT (EMBODIMENT)
Physical/virtual environment, other actors...
▪ Perceived robot actions
▪ Perceived robot appearance
▪ Inferred intentional action
INTERFACE
▪ Listening for sensor data that fit a specific format
ACTION
JOINT ACTION
ACTION
▪ Situation awareness, inferred robot status and predicted next actions
▪ Inferred design metaphor/mental model
▪ Expected physical, perceptual and cognitive capabilities
▪ Perceived agency/intentionality
SITUATIONAL COMMON GROUND
*Needs to be negotiated
▪ "Human-aware", "situation-aware" and "self-aware" in ways specific to its programming
▪ "Beliefs", goal, action plan
▪ Current system model
▪ General world knowledge
COMMON GROUND
▪ Goals
▪ Knowledge base
▪ All situation models

Fig. 3 This figure is based on the architecture in Fig. 1. It is a simplified version of the model that is meant to clarify the model's relation to the design recommendations

state of the device is, making sure that the person has a good conceptual model of how the device operates, and giving feedback in response to the person's actions. Such design principles have also been proposed for HRI [26,88]. Here, we propose several design recommendations based on the AMODAL-HRI model for interaction designers in the field of HRI.

The human and the robot are different types of entities. In the model in Fig. 1, they are depicted in an abstracted form, with similar-yet-different processes and components. This allows for comparing the two sides. Such a comparison can help identify potential communication failures. Differences in the capabilities of the robot versus the human can lead to communication failures if the human has expectations regarding robot behaviour that the robot does not meet. In Sect. 7.1, we discuss some of these differences, such as differences in perceptual capabilities.

In Sects. 7.2–7.6, we discuss design recommendations that are aimed at overcoming or ameliorating these differences. In order to better illustrate the design recommendations and the factors that are most relevant in the communication process, we have added an additional version of the model in Fig. 3. The way the robot processes information is still the same as in the robot architecture version in Fig. 1.

The design recommendations in this section are related to the concept of transparency (and related themes such as explainability, understandability, and interpretability). It has been proposed that transparency can aid a user when it comes to understanding how an AI system works and performs decision-making processes. Transparency for robotics and AI means that the system informs the user regarding what the system is doing and why, (1) making it easier for the end user to predict the system's future actions and (2) in order to enhance the user's trust in the system. The function of transparency is to support the end user in understanding the reasons behind a system's decisions and actions, and it helps the user check if the system is working correctly [33].

7.1 Differences Between the Human and the Robot

Before diving into a discussion of high-level differences between humans and robots, it should be noted that there is an incredible diversity when it comes to human bodies and physical abilities, how humans interact with technologies, a diversity of cognitive abilities, and so forth. Robots should be designed in ways that ensure accessibility in order to accommodate a diverse group of end users. While there is a large variety in robot morphologies and the ways robots can be programmed, robots are far more generic, and the way one robotic system functions can in theory be replicated on (similar) robotic systems.

Robots and humans have different perceptual capabilities, as there is a difference between robot sensors and human sensing. Robot sensors have a different range and may perceive different information, compared to the human sensory organs. For example, the camera view may only cover a fraction of the range and area that a human eye can perceive. Liu and Chai write that humans and robots lack a shared perceptual basis due to differences in perception and reasoning. The robot's perceptual capabilities are limited and markedly different from those of the human interaction partner due to, among other things, its specific computer vision algorithms (one particular problem that is well-researched is the referential grounding problem, which in the context of HRI refers to the problem of connecting references by a human interaction partner to objects in the environment to perceptions of those objects by a robot) [64]. In addition to differences introduced by the particularities of a machine learning algorithm, there are differences in sensor reliability in different conditions [80, p.103], and different modelling conventions can be chosen to represent the world (environment) and the state of the robot based on sensor data [80, p.104]. On the other hand, a robot can have additional sensors, such as infrared sensors, and obtain information a human does not (biologically) have access to. This can lead to the problem that it can be difficult for a human to understand what the robot is (in)capable of doing or perceiving. The robot's capabilities and functionality can be conveyed to the human through training [16] (different instruction methods such as video tutorials are possible, as is trial-and-error exploration [6]). Others propose that the robot could assess the reliability of its perceptions through human–robot dialogue [64].

Another difficulty with respect to robot perceptual and cognitive capabilities concerns asymmetries in the recognition versus production of speech. Thomaz et al. note that it is much more challenging to make robots capable of recognizing speech than it is to make them capable of producing speech with a similar level of complexity. However, if robots are capable of producing speech at a certain level of complexity, this may lead people to infer that the robot will be able to understand their speech at that level of complexity [86].

In addition, robot perception and reasoning can be biased due to biases in the datasets that were used to train machine learning models. For instance, with regards to gender bias, Wang et al. write that in dataset COCO, *"images of plates contain significantly more women than men. If a model predicts that a plate is in the image, we can infer there is likely a woman as well. We refer to this notion as leakage"* [92, p. 5310]. If robots make use of machine learning algorithms that were trained using such datasets, they may amplify and reinforce existing societal biases. While humans are certainly biased as well, our biases are not amplified in a similar way.

With regards to physical movement, humans and robots have different action capabilities: they have a different workspace, and different possibilities regarding motion speed and acceleration. (Again, we note that there are differences among humans as well.) Humans and robots have different action possibilities and communication modalities (for instance, a robot may be able to communicate using LED lights) due to differences in embodiment and morphology. Hoffman et al. write that there is a large variety in robot morphologies. For instance, the robot's embodiment can be zoomorphic or humanoid, some are able to manipulate objects with arms while other have wheels, et cetera. They write that perceived robot morphology influences the capabilities that humans expect a robot to have [48].

It can be expected that humans will adjust more easily to robot limitations than vice versa. Dale et al. [22] describe that in studies of computer-mediated communication, people accommodate to their interaction partner when they think their interaction partner is simulated, for instance by using less complex language and by taking the other's perspective more often, as they aim for a maximum level of mutual understanding. While it may not be desirable to rely on humans to adjust to the limitations of machines as a design strategy, we mention it here as a difference.

7.2 Affordances

Signifying affordances is useful in the context of communicating differences in robot sensing and action capabilities to the human. The term affordance, coined by Gibson, refers to an agent's possibilities for action when interacting with an object or environment [69]. This agent could be a human or a robot. The affordance concept can be applied to a human's possibilities for action when interacting with an object or device such as a robot. In this case, signifying the affordances (through signifiers) in order to make them discoverable is a human-centered design challenge (see [34]). The concepts of affordances and signifiers are closely tied to Norman's concept of discoverability, which is a human-centric notion that indicates that the human can find out how the device functions through directed experimentation or applying mental models from similar devices.

For example, if the robot has the capacity to record audio and interpret speech, the robot affords being spoken to by a human interaction partner. Affordances can be communicated to the human interaction partner by using signifiers. That is, by communicating such things as what the robot can read with its sensors, it can be communicated to the human how the robot can be interacted with. The robot's actions can give the human clues regarding the actions that are possible with the robot. For instance, if this particular robot can only understand the words "yes" or "no", it can signify this by asking the participant to answer with "yes" or "no" after each question.

Another consideration regards embodiment design, for instance with respect to the placement of sensors or how sensor placement is communicated. Humanoid robots are often designed to give the impression of having eyes, but sensors that capture image data are not necessarily placed at the same location. The same goes for microphone and loudspeaker locations. For instance, on the Pepper robot, the microphones are placed on top of the head, while the speakers are at the ear locations [81]. When it comes to signifying affordances, if a humanoid design is chosen, it may be desirable to place sensors in a way that corresponds to the approximate location of human sensory organs. If the sensor placement deviates from expectations significantly, this may need to be communicated to human interaction partners.

Regarding the example in Sect. 6.5, we can identify various opportunities for communicating the state of the teaching and coordination process by applying the concept of affordances. First of all, the robot can indicate that it can be talked to by asking questions regarding objects in the environment or through an attentive posture (affordances/signifiers). Secondly, the robot has already indicated that it can focus its attention and attend to items the human talks about by looking at the object the human is pointing at (see label **Robot (4)**).

Design Recommendation 1 Signify affordances to indicate how the robot can be interacted with.

Design Recommendation 2 Communicate the robot's intended function and action possibilities in different situations to novice users, so the human can understand how the robot functions.

See also Fischer [34], who argues that the robot can communicate affordances implicitly by means of leading questions, and who argues that we can make use of the "downward evidence" signalling strategy: when humans are presented with high-level capabilities, they expect the robot to have lower-level capabilities as well. This is an implicit way of signalling affordances [34].

7.3 Mental Models and Design Metaphors

Mental models have been defined as *"the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states"* [75, p.7]. Interpretation and understanding in humans can be described as occurring through the application of certain frames or mental models [30] to a situation, based on previously experienced situations[6]. In HRI, the mental model concept has been used to indicate people's estimates of the knowledge, abilities, role and goals of the robot [53]. We also link the concept of mental models to the concept of the *design metaphor*. This concept was discussed in the context of HRI by Deng et al. [24], and refers to how the associations that a particular robot design provokes lead people to have certain expectations regarding the way it functions. For instance, if a robot has a humanlike appearance, it appeals to a human design metaphor.

Mental models or design metaphors can also be provoked by means of the robot's behaviour. For instance, Cha et al. [7] report that when a robot has conversational speech abilities, people perceive the robot to have a higher level of physical capabilities than if the robot has functional speech, although this depends on whether the robot is successful at achieving its task. They conclude that functional speech is more effective at setting expectations at an accurate level. Conversational speech, in their experiment, consisted of phatic expressions, while functional speech concerned status information and next actions. This would suggest that conversational speech evokes expectations of social agency, while functional speech helps set expectations more correctly, as it is more in line with a device mental model or device design metaphor.

Thus, appealing to specific design metaphors or mental models may help people build correct expectations of how the robot functions. We do note that designers should take care not to reinforce societal stereotypes (e.g. regarding gender) when choosing to appeal to, for instance, human design metaphors.

Design Recommendation 3 Use an appropriate design metaphor to set human expectations of the robot and the interaction at a more accurate level.

[6] Endsley relates the concept of the mental model to that of the situation(al) model, a way of understanding the current state of a system. A situation model/mental model can be used to identify critical cues and elements to attend to, understand the meaning of elements in a situation, predict future states, and identify which actions are appropriate in this situation. Mental models and schemata are based on experience [30]. Similar terms have been proposed by different theories in communication research. For instance, similar terms are *recipe knowledge* [5, p.57] and *habitualized actions* [5, p.71] in social constructionism (the sociocultural tradition in communication theory). In action-assembly theory, a related concept is *procedural knowledge*, which helps an individual to determine what to say or do next (sociopsychological tradition of communication theory). In Goffman's frame analysis, we find the terms *strips* and *frames* [63] (sociocultural tradition). Goffman set forth the notion of *primary frameworks* that, he says, individuals use as *"schemata of interpretation"* [38, p.21].

7.4 Transparency

Another design recommendation is to make use of transparency mechanisms. Transparency involves communicating such things as the robot state, (accuracy of) sensing capabilities and currently active processes within the robot. By communicating the robot's limitations and internal processes, humans can form a more helpful mental model or perform behaviours that accommodate the robot's limitations [71]. Fischer's notion of transparency involves communicating reasons for robot failure, the robot's reliability, and robot awareness of the human to the human interaction partner [34]. The concept is related to Johnson et al.'s notion of *observability* [52], as well as *understandability*: making sure the robot's behaviour is understandable to its human interaction partner by making such things as the robot's beliefs and goals [44]. Transparency through visualization has been investigated in the context of robotics [9,71,79]. Communicating speech recognition results is an example of transparency regarding the sensing capabilities of the robot. A less taxing alternative with respect to human attention may be to indicate the accuracy of its speech recognition. Indicating whether the robot is currently processing input by changing the colour of a subset of its LEDs is an example of transparency regarding internal robot processes.

Regarding the example in Sect. 6.5, the success or failure in sensing the human's actions can be indicated by mechanisms that enhance system transparency. For instance, (1) the robot can indicate that the human's speech was not understood by verbally informing the human or by providing speech recognition results on a screen. (2) The robot may also provide transparency with respect to common ground by indicating which objects were recognized (verbally, or on a screen).

Design Recommendation 4 Use system transparency to communicate status information, sensing capabilities and currently active processes.

We note that finding the right level of transparency is a non-trivial task. A variety of communication modalities can be used to different effects, and other factors such as cognitive overload may start to play a role when a great deal of information is communicated to an end user. Testing the effects of different ways of conveying information as well as different types of information can be a labour-intensive process.

7.5 External Influence

One difference between humans and robots is that external influences on the human side require interpretation by the human to have an effect on the human (with the exception of direct physical impacts), while external influence on the robot side is always direct. Humans can be directly influenced by impacting which information reaches the human or manipulating their body (e.g. turning their head to face in a certain direction). On the other hand, the robot has an 'open' nature; it is permeable to external influences (provided it is reprogrammable and reconfigurable), see also Fig. 2. External influences and connections are not always problematic, but a few cases require further consideration. External influences and external data processing should be communicated to end users, especially in cases in which the end user's privacy is impacted. The external influences should be made explicit for the end user, and the end user should be asked for consent regarding external data processing. One can also think of the case of software updates. If it is important for the robot to maintain functionality even in the absence of a stable internet connection, the system developers should build a version of the system that still provides the desired functionality without external processing.

Design Recommendation 5 Communicate external influences to end users and, if possible and necessary, supply a product that is still functional without external processing.

Again, this problem is non-trivial, as regulations such as the GDPR need to be taken into account, as well as rights such as the right to privacy. The European GDPR regulation requires companies to provide end users with intelligible explanations regarding the way their data is used [31], which is also related to the issue of transparency as described in Sect. 7.4. Felzmann et al. provide considerations regarding robots, the GDPR and transparency, and propose a procedural checklist for implementing transparency within robotics development. The checklist includes steps such as identifying obligations, as well as stakeholders and their needs [33]. We also note that privacy and data processing must be even more carefully considered when it comes to social robots operating in public space. While guidelines for video surveillance with static cameras have been developed in the EU [28], for instance, social robots that move around in space autonomously and are equipped with cameras would require more specific guidelines and regulations.

7.6 Common Ground

Achieving mutual predictability would require that the robot shares representations with a human interaction partner and 'understands' them in a similar way. Beliefs held by both entities can be considered common ground, although these beliefs are present in different ways in the human and the robot agent. For instance, on the robot side, beliefs can be stored in the form of logical statements. Note that with respect to common ground in the example in Sect. 6.5, *"cup"*, *'cup'*, **[object1]** and the actual cup are all different things.

With respect to instrumental uncertainty, common ground uncertainty and motivational uncertainty, interaction designers can choose to design robot behaviour in a way that reduces uncertainty of a human interaction partner. With respect to the example in Sect. 6.5, the robot could verbally communicate its goal at the start of the interaction to reduce motivational uncertainty.

Design Recommendation 6 Integrate specific robot behaviours to reduce instrumental, motivational and common ground uncertainty.

Solutions for disconnects in common ground include making robot capabilities explicit, e.g. by verbally or textually informing a user (system transparency). Another mechanism is to include external representations of the joint activity. This has been discussed by Clark, who gives the example of the chess board, a device that keeps track of the joint activity, that is, chess [14, p.45]. In HRI, screens and user interfaces can play the role of such an external representation. In the manufacturing domain, collaborative robot systems often include a graphical user interface that displays status information and task progress.

Design Recommendation 7 Use external representations of the joint activity to keep track of the accumulated common ground, if necessary.

Note that the 'common ground' indicated in the model will always be minimal compared to the common ground shared by humans. For instance, even if two humans cannot speak each other's language, they can oftentimes still communicate and understand each other. If a person encounters a robot that cannot interpret the language (or way of interacting) the human uses, the interaction will completely fail. In interaction between humans, we can assume a substantial common ground, which cannot be assumed in human–robot interactions [16].

7.7 Recommendations for Modelling Human–robot Interactions

Based on the models surveyed in this paper and the process of modelling that led to the models proposed in this paper, we would like to give HRI researchers some recommendations with regards to modelling communication and interaction processes in HRI.

Model Design Recommendation 1 Define the level of analysis when discussing and modelling human–robot interactions or communication between humans and robots. It is not possible to include every level of discussion, nor every relevant factor, when modelling an interaction.

Model Design Recommendation 2 Clarify design choices and consider the assumptions made in choosing to model the interaction in a certain way. Be specific in presenting the supposed functioning of the human, robot and interaction. Make it explicit and keep in mind that there is a large variety of human bodies, abilities, behaviours and identities.

8 Limitations of the Present Work

There are limits to applying models of communication between humans to communication between humans and robots. However, we posit that models of communication between humans are a useful starting point, as they allows us to directly compare similar processes in communication between humans to communication between humans and robots, and humans likely bring expectations from communication between humans to their interactions with robots. This can give us insight into when, why and how communication failures may arise.

The model by Kincaid is not the only model of human perception, cognition and action (see e.g. [61]). Human cognition can most likely be depicted in a more accurate way, but this was not the aim of the current paper. The model by Kincaid was chosen as it meshes well with a joint action perspective, which is useful in the context of HRI and HRC. We hope we have given sufficient background to demonstrate that other types of models and research on communication theory can also be applied. We have outlined connections to different levels of discussion in communication theory. Here, we see another potential area of future research: models on the level of group communication and interaction, as well as on the level of organizations and society. At the group (or even media) level, one-to-many and many-to-many types of interactions can be considered.

Timing is of high importance in joint actions. The models in Figs. 1 and 2 are depicted as processes, but do not detail exactly *when* communication is necessary. Changes in timing have an effect on how the action is interpreted. We have taken some initial steps towards incorporating the time dimension in the example in Sect. 6.5, but additional models and frameworks may be necessary. The model by Hellström and Bensch proposes that communication is necessary when one agent (agent X) determines there is a mismatch between the other agent Y's *estimation of agent X's state* and *agent X's actual state* [44]. This can be derived for the example in Sect. 6.5 in a similar way. However, the time dimension is of such importance that it deserves more prominence in a model of interaction. We would therefore also encourage other researchers to explore alternatives and come up with proposals that better express embodied, spatio-temporal and contextual aspects.

One question that arose during this work was the question of whether we can speak of communication at all when it comes to HRI, as using intentionalist vocabulary to describe robot behaviour may be too suggestive of human-level capabilities. However, as technologies advance, people may attribute communication capabilities to the robotic system anyway. This means that there are ethical consequences associated with this question, for instance regarding deception [18].

In this article, we connect the literature on HRI to (mainly) the broad field of communication theory. Our work has been especially influenced by symbolic interactionism and social constructionism [5]. Other influences are cognitive psychology and psycholinguistics (joint action), and theory on situation awareness [30]. We are aware that our thinking is highly influenced by tendencies common in European thought and research traditions (rational, focused on intentionality, cognitive, individual). For instance, it can be observed that the proposed model and the design recommendations place a large emphasis on cognitive processes and understanding. We invite other researchers to criticize perceived gaps in the arguments presented here and to propose alternatives.

9 Conclusion

The first aim of this article was to connect the research field of HRI to that of communication theory. We surveyed models of interpersonal communication from communication theory and focused our discussion on the transmission model of communication and transactional models of communication. We discussed communication and interaction models that are presently applied in HRI. We identified several models that fit a control paradigm of human–robot interactions, and models that fit a social interaction paradigm. We identified and discussed several problematic aspects of existing communication and interaction models in HRI. The main problem we identified, is that often, the human and the robot are depicted as similar entities, while they clearly are dissimilar at the moment. This was in line with our second aim: to identify the asymmetries in human–robot interaction and communication. Differences in capabilities do not have to be problematic, as the robot's capabilities can be complementary to those of a human. However, communication failures as a result of these differences may arise. Another problem is that the interaction itself is often depicted in a simplified way, and understood as the 'sending of signals'. A joint action approach is more appropriate. The third aim of this article was to formalize an asymmetric model of joint action for HRI. We proposed the Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI). We did not aim to make the model as asymmetric as possible; instead, we aimed for the model to have similar processes on the human and the robot side to allow for direct comparison. This allows for identifying differences in a productive way: it allows for identifying asymmetries between human and robot capabilities and for proposing strategies to improve the robot's usability with respect to said asymmetries. In terms of practical applications, the model can be adapted to fit a specific technical setup. We demonstrated how the general model can be useful in practice, namely by means of the use of scripts as in the example in Sect. 6.5 and by comparing components and critically discussing the results of the comparison and differences with interpersonal interaction.

The main contribution of this work regards improving human mental models of robots, by investigating how interaction design can contribute to improving people's mental models of robots and their capabilities, in order to achieve successful human–robot interactions. By using this concept, we assume that people's previous experiences with technologies, objects, and even humans impact their expectations of interactions with devices such as robots. People's expectations can change by learning about or repeatedly interacting with the technology. We assume that if we achieve a better match between expected and actual robot behaviour, we will foster social acceptance and trust. Supporting accurate understanding of systems will help people know how to use the technology for their own goals, and help people rely on technology appropriately [60].

10 Future Work

As mentioned earlier, some aspects of the proposed model deserve more attention, such as timing in interaction, aspects relating to the environment (such as embodiment, physical space), and the involvement of other actors and team or group coordination. Future work can include surveying coordination frameworks and cognitive architectures for coordination, as well as approaches that model timing in interaction, with the aim of proposing additional communication and interaction models. We may also propose additional models that operate on different levels of communication (e.g. the level of group interaction, of organizations, the media, and society). Another interesting approach would be to adapt the model so that it integrates an existing cognitive architecture, for instance one based on a three-tiered architecture. This can be useful to see if the model still applies or breaks down. Finally, we propose that more work is required to detail how we can design transparent user interfaces for HRI applications.

Acknowledgements This work was supported by Doctoral College TrustRobots, Technische Universität Wien. We would like to thank all anonymous reviewers, Michael Koller, and Guglielmo Papagni for their feedback on previous versions of this paper.

Funding Open access funding provided by TU Wien (TUW).

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Alač M (2016) Social robots: things or agents? AI Soc 31(4):519–535. https://doi.org/10.1007/s00146-015-0631-6
2. Barnlund DC (1970) A transactional model of communication. Language behavior: a book of readings in communication. Mouton, The Hague, pp 43–61
3. Bauer A, Wollherr D, Buss M (2008) Human–robot collaboration: a survey. Int J Humanoid Rob 5(1):47–66. https://doi.org/10.1142/S0219843608001303
4. Bensch S, Jevtić A, Hellström T (2017) On interaction quality in human–robot interaction. In: Proceedings of the 9th international conference on agents and artificial intelligence, SCITEPRESS. Science and Technology Publications, Porto, Portugal, pp 182–189. https://doi.org/10.5220/0006191601820189
5. Berger PL, Luckmann T (1966) The social construction of reality: a treatise in the sociology of knowledge. Penguin Books
6. Cakmak M, Takayama L (2014) Teaching people how to teach robots: the effect of instructional materials and dialog design. In: 2014 9th ACM/IEEE international conference on human–robot interaction (HRI), pp 431–438. https://doi.org/10.1145/2559636.2559675
7. Cha E, Dragan AD, Srinivasa SS (2015) Perceived robot capability. In: 2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 541–548. https://doi.org/10.1109/ROMAN.2015.7333656
8. Chandler D (1994) The transmission model of communication. http://visual-memory.co.uk/daniel/Documents/short/trans.html
9. Chen JY, Procci K, Boyce M, Wright J, Garcia A, Barnes M (2014) Situation awareness-based agent transparency. Tech. rep., Defense technical information center, fort belvoir, VA. https://doi.org/10.21236/ADA600351
10. Chong HQ, Tan AH, Ng GW (2007) Integrated cognitive architectures: a survey. Artif Intell Rev 28(2):103–130. https://doi.org/10.1007/s10462-009-9094-9
11. Chung H, Iorga M, Voas J, Lee S (2017) Alexa, can i trust you? Computer 50(9):100–104. https://doi.org/10.1109/MC.2017.3571053
12. Clark HH (1996a) Common ground. In: Using language. Cambridge University Press
13. Clark HH (1996b) Joint actions. In: Using language. Cambridge University Press
14. Clark HH (1996c) Joint activities. In: Using language. Cambridge University Press
15. Clark HH (1996d) Meaning and understanding. In: Using Language. Cambridge University Press
16. Clodic A, Pacherie E, Alami R, Chatila R (2017) Key elements for human–robot joint action. In: Hakli R, Seibt J (eds) Sociality and normativity for robots. Springer, pp 159–177. https://doi.org/10.1007/978-3-319-53133-5_8
17. Coeckelbergh M (2011) You, robot: on the linguistic construction of artificial others. AI Soc 26(1):61–69. https://doi.org/10.1007/s00146-010-0289-z
18. Coeckelbergh M (2018) How to describe and evaluate "deception" phenomena: recasting the metaphysics, ethics, and politics of ICTs in terms of magic and performance and taking a relational and narrative turn. Ethics Inf Technol 20(2):71–85. https://doi.org/10.1007/s10676-017-9441-5
19. Cohen PR, Levesque HJ (1991) Teamwork. Nous 25:487–512
20. Craig RT (1999) Communication theory as a field. Commun Theory 9(2):119–161. https://doi.org/10.1111/j.1468-2885.1999.tb00355.x
21. Curioni A, Knoblich G, Sebanz N (2017) Joint action in humans: a model for human–robot interactions. In: Goswami A, Vadakkepat P (eds) Humanoid robotics: a reference. Springer Netherlands, pp 1–19. https://doi.org/10.1007/978-94-007-7194-9_126-1
22. Dale R, Fusaroli R, Duran ND, Richardson DC (2014) The Self-Organization of Human Interaction. In: The psychology of learning and motivation. Elsevier Inc.: Academic Press, pp 43–96. https://doi.org/10.1016/B978-0-12-407187-2.00002-2
23. Dautenhahn K (2007) Socially intelligent robots: dimensions of human–robot interaction. Philos Trans R Soc B: Biol Sci 362(1480):679–704. https://doi.org/10.1098/rstb.2006.2004
24. Deng E, Mutlu B, Mataric MJ (2019) Embodiment in socially interactive robots. Found Trends Robot 7(4):251–356. https://doi.org/10.1561/2300000056
25. Dragan AD, Bauman S, Forlizzi J, Srinivasa SS (2015) Effects of robot motion on human–robot collaboration. In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction (HRI '15). ACM Press, pp 51–58. https://doi.org/10.1145/2696454.2696473
26. Drury JL, Hestand D, Yanco HA, Scholtz J (2004) Design guidelines for improved human–robot interaction. In: Extended abstracts of the 2004 conference on Human factors and computing systems (CHI '04). ACM Press, Vienna, Austria, p 1540. https://doi.org/10.1145/985921.986116
27. Dumas B, Lalanne D, Oviatt S (2009) Multimodal interfaces: A survey of principles, models and frameworks. In: Lalanne D, Kohlas J (eds) Human Machine Interaction, vol 5440. Springer, Berlin. pp 3–26. https://doi.org/10.1007/978-3-642-00437-7_1
28. (EDPS) EDPS (2010) The EDPS video-surveillance guidelines. https://edps.europa.eu/sites/edp/files/publication/10-03-17_video-surveillance_guidelines_en.pdf
29. Emmanouil TA, Ro T (2014) Amodal completion of unconsciously presented objects. Springer Psychon Bull Rev. https://doi.org/10.3758/s13423-014-0590-9
30. Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. Human Factors: J Human Factors Ergon Soc 37(1):32–64. https://doi.org/10.1518/001872095779049543
31. EUR-Lex Access to European Union Law POotEU (2016) Regulation (EU) 2016/679 of the European Parliament and of the

Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504

32. Feine J, Gnewuch U, Morana S, Maedche A (2019) A taxonomy of social cues for conversational agents. Int J Hum Comput Stud 132:138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009
33. Felzmann H, Fosch-Villaronga E, Lutz C, Tamo-Larrieux A (2019) Robots and transparency: the multiple dimensions of transparency in the context of robot technologies. IEEE Robot Autom Mag 26(2):71–78. https://doi.org/10.1109/MRA.2019.2904644
34. Fischer K (2017) When transparent does not mean explainable. In: Proceedings of 'Explainable Robotic Systems', workshop in conjunction with the HRI 2018 conference, Chicago, p 3
35. Fong T, Thorpe C, Baur C (2002) Robot as partner: vehicle teleoperation with collaborative control. In: Schultz AC, Parker LE (eds) Multi-robot systems: from swarms to intelligent automata. Springer Netherlands, Dordrecht. pp 195–202. https://doi.org/10.1007/978-94-017-2376-3_21
36. Forlizzi J (2008) The product ecology: understanding social product use and supporting design culture. Int J Des 2(1):11–20
37. Fuchs C (2017) Social media, a critical introduction, 2nd edn. SAGE Publications
38. Goffman E (1974) Primary frameworks. In: Frame analysis—an essay on the organization of experience. Northeastern University Press, pp 21–39
39. Goodrich MA, Schultz AC (2007) Human–robot interaction: a survey. Foundations and trends®in human–computer interaction 1(3):203–275. https://doi.org/10.1561/1100000005
40. Gunkel DJ (2012) Communication and artificial intelligence: opportunities and challenges for the 21st century. Communication +1 1(1):26. https://doi.org/10.7275/R5QJ7F7R
41. Guzman AL, Lewis SC (2020) Artificial intelligence and communication: a human–machine communication research agenda. New Media Soc 22(1):70–86. https://doi.org/10.1177/1461444819858691
42. Hassenzahl M, Borchers J, Boll S, Rosenthal vd Pütten AM, Wulf V, (2021) Otherware: how to best interact with autonomous systems. Interactions. https://doi.org/10.1145/3436942
43. Hegel F, Gieselmann S, Peters A, Holthaus P, Wrede B (2011) Towards a typology of meaningful signals and cues in social robotics. In: 2011 RO-MAN, IEEE, Atlanta, GA, USA, pp 72–78. https://doi.org/10.1109/ROMAN.2011.6005246
44. Hellström T, Bensch S (2018) Understandable robots-what, why, and how. Paladyn J Behav Robot 9(1):110–123. https://doi.org/10.1515/pjbr-2018-0009
45. Hiatt LM, Narber C, Bekele E, Khemlani SS, Trafton JG (2017) Human modeling for human-robot collaboration. Int J Robot Res 36(5–7):580–596. https://doi.org/10.1177/0278364917690592
46. Hoffman G (2019) Evaluating fluency in human–robot collaboration. IEEE Trans Human–Machine Syst 49(3):209–218. https://doi.org/10.1109/THMS.2019.2904558
47. Hoffman G, Breazeal C (2004) Collaboration in human–robot teams. In: AIAA 1st intelligent systems technical conference, American Institute of Aeronautics and Astronautics, Chicago, Illinois. https://doi.org/10.2514/6.2004-6434
48. Hoffmann L, Bock N, Rosenthal vd Pütten AM (2018) The peculiarities of robot embodiment (EmCorp-scale): development, validation and initial test of the embodiment and corporeality of artificial agents scale. In: Proceedings of the 2018 ACM/IEEE international conference on human–robot interaction. ACM, pp 370–378. https://doi.org/10.1145/3171221.3171242
49. Hurley S (2008) The shared circuits model (SCM): how control, mirroring, and simulation can enable imitation, deliberation, and mindreading. Behav Brain Sci 31(1):1–22. https://doi.org/10.1017/S0140525X07003123
50. Ihde D (1990) Technology and the Lifeworld. Indiana University Press, From Garden To Earth
51. Ihde D (2009) Postphenomenology and technoscience, the Peking University lectures. Suny Press, Albany
52. Johnson M, Bradshaw JM, Feltovich PJ, Jonker CM, Van Riemsdijk MB, Sierhuis M (2014) Coactive design: designing support for interdependence in joint activity. J Human–Robot Interact 3(1):43. https://doi.org/10.5898/JHRI.3.1.Johnson
53. Kiesler S (2005) Fostering common ground in human–robot interaction. In: ROMAN 2005. IEEE international workshop on robot and human interactive communication, 2005, pp 729–734. https://doi.org/10.1109/ROMAN.2005.1513866
54. Kincaid DL (1979) The convergence model of communication. Papers of the East-West Communication Institute No. 18:52
55. Kotseruba I, Tsotsos JK (2020) 40 years of cognitive architectures: core cognitive abilities and practical applications. Artif Intell Rev 53(1):17–94. https://doi.org/10.1007/s10462-018-9646-y
56. Kruijff GJM (2013) Symbol grounding as social, situated construction of meaning in human–robot interaction. KI - Künstliche Intelligenz 27(2):153–160. https://doi.org/10.1007/s13218-013-0238-3
57. Kruse T, Pandey AK, Alami R, Kirsch A (2013) Human-aware robot navigation: a survey. Robot Auton Syst 61(12):1726–1743. https://doi.org/10.1016/j.robot.2013.05.007
58. Krämer NC, von der Pütten A, Eimler S (2012) Human-agent and human–robot interaction theory: Similarities to and differences from human–human interaction. In: Zacarias M, de Oliveira JV (eds) Human–computer interaction: the agency perspective, vol 396. Springer, Berlin, pp 215–240. https://doi.org/10.1007/978-3-642-25691-2_9
59. Lackey SJ, Barber DJ, Martinez SG (2014) Recommended considerations for human–robot interaction communication requirements. In: Kurosu M (ed) Human–computer interaction. Advanced interaction modalities and techniques, vol 8511, Springer, pp 663–674. https://doi.org/10.1007/978-3-319-07230-2_63
60. Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Human Factors: J Human Factors Ergon Soc 46(1):50–80. https://doi.org/10.1518/hfes.46.1.50_30392
61. Lemerise EA, Arsenio WF (2000) An integrated model of emotion processes and cognition in social information processing. Child Dev 71(1):107–118. https://doi.org/10.1111/1467-8624.00124
62. Lindesmith AR, Strauss AL, Denzin NK (1999) Social psychology, 8th edn. SAGE Publications
63. Littlejohn SW, Foss KA (2011) Theories of human communication, 10th edn. Waveland Press
64. Liu C, Chai JY (2015) Learning to mediate perceptual differences in situated human–robot dialogue. In: Proceedings of the twenty-ninth AAAI conference on artificial intelligence, pp 2288–2294
65. Lorenz T, Weiss A, Hirche S (2016) Synchrony and reciprocity: key mechanisms for social companion robots in therapy and care. Int J Social Robot 8(1):125–143. https://doi.org/10.1007/s12369-015-0325-8
66. Malik AA, Bilberg A (2019) Developing a reference model for human–robot interaction. Int J Interact Des Manuf (IJIDeM). https://doi.org/10.1007/s12008-019-00591-6
67. Mirnig N, Weiss A, Tscheligi M (2011) A communication structure for human–robot itinerary requests. In: 2011 6th ACM/IEEE international conference on human–robot interaction (HRI), pp 205–206. https://doi.org/10.1145/1957656.1957733
68. Mutlu B, Terrell A, Huang CM (2013) Coordination mechanisms in human—robot collaboration. In: Proceedings of the HRI 2013 workshop on collaborative manipulation, p 6
69. Norman DA (2013) The design of everyday things, revised and expanded. Basic Books, New York

70. Pack AA (2018) Language research: dolphins. In: Vonk J, Shackelford T (eds) Encyclopedia of animal cognition and behavior. Springer, Berlin, pp 1–10
71. Perlmutter L, Kernfeld E, Cakmak M (2016) Situated language understanding with human-like and visualization-based transparency. In: Robotics: science and systems XII, robotics: science and systems foundation. https://doi.org/10.15607/RSS.2016.XII.040
72. Pickering MJ, Garrod S (2013) An integrated theory of language production and comprehension. Behav Brain Sci 36(4):329–347. https://doi.org/10.1017/S0140525X12001495
73. Reeves B, Nass C (1996) The media equation. how people treat computers, television, and new media like real people and places. Cambridge University Press, pp 3–15
74. Rios-Martinez J, Spalanzani A, Laugier C (2015) From proxemics theory to socially-aware navigation: a survey. Int J Soc Robot 7(2):137–153. https://doi.org/10.1007/s12369-014-0251-1
75. Rouse WB, Morris NM (1985) On looking into the black box: prospects and limits in the search for mental models. Technical report. Center for Man–Machine Systems Research, School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta GA 30332
76. Sandry E (2015) Robots and communication. Palgrave Macmillan, Palgrave pivot
77. Seibt J (2018) Classifying forms and modes of co-working in the ontology of asymmetric social interactions (OASIS). In: Frontiers in artificial intelligence and applications, pp 133–146. https://doi.org/10.3233/978-1-61499-931-7-133
78. Shannon C, Weaver W (1964) The mathematical theory of communication, first paperbound edition, tenth, printing. The University of Illinois Press, Urbana
79. Sibirtseva E, Kontogiorgos D, Nykvist O, Karaoguz H, Leite I, Gustafson J, Kragic D (2018) A comparison of visualisation methods for disambiguating verbal requests in human–robot interaction. In: 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 43–50. https://doi.org/10.1109/ROMAN.2018.8525554
80. Siciliano B, Khatib O (2008) Springer handbook of robotics. Springer, Berlin. https://doi.org/10.1007/978-3-540-30301-5
81. Softbank Robotics (2017) Technical overview—aldebaran 2.5.11.14a documentation. http://doc.aldebaran.com/2-5/family/pepper_technical/index_pep.html, last visited on 2020-08-07
82. Spiel K (2021) The bodies of TEI—investigating norms and assumptions in the design of embodied interaction. In: TEI ’21: Proceedings of the fifteenth international conference on tangible, embedded, and embodied interaction, pp 1–19. https://doi.org/10.1145/3430524.3440651
83. Sung J, Grinter RE, Christensen HI (2010) Domestic robot ecology: an initial framework to unpack long-term acceptance of robots at home. Int J Soc Robot 2(4):417–429. https://doi.org/10.1007/s12369-010-0065-8
84. Taniguchi T, Ugur E, Hoffmann M, Jamone L, Nagai T, Rosman B, Matsuka T, Iwahashi N, Oztop E, Piater J, Wörgötter F (2019) Symbol emergence in cognitive developmental systems: a survey. IEEE Trans Cognit Dev Syst 11(4):494–516. https://doi.org/10.1109/TCDS.2018.2867772
85. Tenorth M, Beetz M (2017) Representations for robot knowledge in the KnowRob framework. Artif Intell 247:151–169. https://doi.org/10.1016/j.artint.2015.05.010
86. Thomaz A, Hoffman G, Cakmak M (2016) Computational human–robot interaction. Found Trends Robot 4(2):104–223. https://doi.org/10.1561/2300000049
87. Thomaz AL, Berlin M, Breazeal C (2005) An embodied computational model of social referencing. In: ROMAN 2005. In: IEEE international workshop on robot and human interactive communication, 2005, pp 591–598. https://doi.org/10.1109/ROMAN.2005.1513844
88. Tsui KM, Abu-Zahra K, Casipe R, M’Sadoques J, Drury JL (2010) Developing heuristics for assistive robotics. In: 2010 5th ACM/IEEE international conference on human–robot interaction (HRI), pp 193–194. https://doi.org/10.1109/HRI.2010.5453198
89. Van Camp J (2019) My jibo is dying and it’s breaking my heart. https://www.wired.com/story/jibo-is-dying-eulogy/
90. Vesper C, Abramova E, Bütepage J, Ciardo F, Crossey B, Effenberg A, Hristova D, Karlinsky A, McEllin L, Nijssen SRR, Schmitz L, Wahn B (2017) Joint action: mental representations, shared information and general mechanisms for coordinating with others. Front Psychol. https://doi.org/10.3389/fpsyg.2016.02039
91. de Visser EJ, Peeters MMM, Jung MF, Kohn S, Shaw TH, Pak R, Neerincx MA (2019) Towards a theory of longitudinal trust calibration in human–robot teams. Int J Soc Robot. https://doi.org/10.1007/s12369-019-00596-x
92. Wang T, Zhao J, Yatskar M, Chang KW, Ordonez V (2019) Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. In: 2019 IEEE/CVF international conference on computer vision (ICCV). IEEE, pp 5309–5318. https://doi.org/10.1109/ICCV.2019.00541
93. William Evans A, Marge M, Stump E, Warnell G, Conroy J, Summers-Stay D, Baran D (2017) The future of human robot teams in the army: factors affecting a model of human-system dialogue towards greater team collaboration. In: Savage-Knepshield P, Chen J (eds) Advances in human factors in robots and unmanned systems, vol 499. Springer, pp 197–209. https://doi.org/10.1007/978-3-319-41959-6_17
94. Yakin HSM, Totu A (2014) The semiotic perspectives of Peirce and Saussure: a brief comparative study. Proc Soc Behav Sci 155:4–8. https://doi.org/10.1016/j.sbspro.2014.10.247
95. Yan H, Ang MH, Poo AN (2014) A survey on perception methods for human–robot interaction in social robots. Int J Social Robot 6(1):85–119. https://doi.org/10.1007/s12369-013-0199-6
96. Yanco HA, Drury J (2004) Classifying human–robot interaction: an updated taxonomy. In: 2004 IEEE international conference on systems, man and cybernetics, vol 3, pp 2841–2846. https://doi.org/10.1109/ICSMC.2004.1400763
97. Young JE, Sung J, Voida A, Sharlin E, Igarashi T, Christensen HI, Grinter RE (2011) Evaluating human–robot interaction: focusing on the holistic interaction experience. Int J Social Robot 3(1):53–67. https://doi.org/10.1007/s12369-010-0081-8
98. Zafari S, Koeszegi ST (2020) Attitudes toward attributed agency: role of perceived control. Int J Social Robot. https://doi.org/10.1007/s12369-020-00672-7

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.