Embodied Cooperative Systems: From Tool to Partnership

Ipke Wachsmuth
Bielefeld University, Germany
e-mail: ipke.wachsmuth@uni-bielefeld.de

Abstract Understanding others' intentions and representing them as being able to understand intentions are relevant factors in cooperation, as is the ability to represent shared goals and coordinated action plans (joint intentions). To endow artificial systems with cooperative functionality, they need to be enabled to adopt the goals of another individual and act together with the other to achieve these goals. Such systems may be embodied as robotic agents or as humanoid agents projected in virtual reality ("embodied cooperative systems"). A central question is how the processes involved interact and how their interplay can be modeled. For example, inter-agent cooperation relies very much on common ground, i.e. the mutually shared knowledge of the interlocutors. Nonverbal behaviors such as gaze and gestures are important means of coordinating attention between interlocutors (joint attention) in the pursuit of goals. In the context of cooperative settings, the view that humans are users of a certain "tool" has shifted to that of a "partnership" with artificial agents, insofar they can be considered as being able to take initiative as autonomous entities. This chapter will outline these ideas taking the virtual humanoid agent "Max" as an example.

Keywords: cooperation; intentions; joint intention; theory of mind; joint attention; artificial agents; BDI

1 Introduction

The idea of embodied cooperative systems pursues a vision of systems that are helpful to humans by making interaction between humans and artificial systems natural and efficient. The long-term objective of our research is a thorough understanding of the processes and functional constituents of cognitive interaction in order to replicate them

in artificial systems that can communicate and cooperate with humans in a natural way. While "cooperative system" could be said to mean a pair (or group) of individuals acting together in the attempt to accomplish a common goal, we prefer the notion of a system exhibiting cooperative behavior by taking on (some of) the goals of another individual and acting together with the other to achieve these shared goals. Cooperation thus involves, as we shall explain further below, some kind of joint intention, which means the ability to represent coordinated action plans for shared goals. Crucial for such cooperation is communication, and when we speak of embodied systems here, the idea is that these systems "by nature" can also employ nonverbal behaviors in cooperative dialogue when coordinating actions between agents.

This contribution is written from the perspective of artificial intelligence which, as an academic discipline, is concerned with building machines – artificial agents – that model human intelligent behaviors and exploit them in technical applications. Such behaviors typically include the functions of perceiving, reasoning, and acting. Research has been moving on towards envisioning artificial agents (e.g. autonomous robots) as partners rather than tools with whom working 'shoulder-by-shoulder' with humans can be effective (Breazeal et al. 2004). Then such systems will also need to incorporate capacities which enable them to align with their human interactant through shared beliefs and intentions. When we thus view agents as intentional systems, a central idea is that their behavior can be understood by attributing them beliefs, desires and intentions (see Sect. 2). The questions particularly addressed from this perspective in this chapter are the following:

1. How can joint intentions and cooperation be modeled and simulated?
2. Can we attribute joint intention to a system or team involving both, a human and an artificial agent?

One of the most basic mental skills is inferring intentions – the ability to see others as intentional agents and to understand what someone else is doing. Intentions are not directly observable, thus they need to be inferred from the interactant's overt behaviors. The types of information exploited to infer intentions are comprised by the interactant's verbal behavior, gaze and facial expression, gestures, as well as the perceived situation and prior knowledge. Hence inferring intentions is not a monolithic

mental faculty, but a composite of different mechanisms including attentive processes (i.e. processes enabling a system to actively focus on a target) and more general cognitive processes such as memory storage or reasoning. Both, understanding others' intentions and representing them as being able to understand intentions, are relevant factors in cooperation.

Human beings (and certain animals) can develop a mental representation of the other, making assumptions (possibly false ones) about the other's beliefs, desires, intentions and probable actions – a 'Theory of Mind' (Premack and Woodruff 1978). Theory of Mind (ToM) refers to the ability to understand others as rational agents, whose behavior is lead by intentional states like beliefs and desires. There are two aspects of Theory of Mind: 'cognitive' ToM (inferring intentional states of the other), and 'affective' ToM (inferring emotional states of the other), referring to an understanding of what the other is likely to do resp. what are the other's feelings. These ideas may also be a valuable prerequisite in modeling communication with virtual humans (Krämer 2008). While our own work has included artificial agents that can infer another agent's emotional state (e.g. Boukricha et al. 2011), we shall in the following mainly focus on the intentional states involved in cooperation.

To endow artificial systems with cooperative functionality, they need to be enabled to adopt (some of) the goals of another individual and act together with the other to achieve these shared goals.[1] Acting together requires that intentional agents engage with one another to form a joint intention, i.e. represent coordinated action plans for achieving their common goals in joint cooperative activity (Tomasello et al. 2005; Bratman 1992). The activity itself may be more simple (e.g. engaging in a conversation; see Sect. 2) or complex (like constructing a model airplane; see Sect. 3). For collaborative engagement, Tomasello et al. (2005) further stress the importance of *joint attention* (mutual knowledge of interactants sharing their focus of attention; see Sect. 4) as well as interactants' ability to reverse action roles and help the other if needed.

If we want to construct artificial systems that are helpful to humans – interacting with us like "partners" –, then such systems should be able to understand and respond to the human's wants in order to be assistive in a given situation. Technically, this

[1] The aspect of "mutual benefit" often included in definitions of cooperation is not taken up here because artificial systems do not seem to have genuine interests that could benefit from cooperation; see (Stephan et al. 2008).

challenge involves the implementation of a range of skills such as: processing language, gaze and gestures, representing intentional states for self and other, detecting and manipulating the other's attention, responding to bids for joint attention, accomplishing goals in joint activity.

In the remainder of this chapter, we will outline these ideas taking the virtual humanoid agent "Max" as a model for a communicative and cooperative agent. We will first look at Max as a "conversational machine" and describe its cognitive architecture, then move on to cooperation by examining details of a cooperative construction scenario, focus on the coordination of attention, and conclude with a view of how the perception of artificial systems may change from tool to partnership.

2 Conversational Machines

The development of conversational machines, i.e. machines that can conduct human-like dialogue, has been a goal of artificial intelligence research for long (Schank 1971; Cassell et al. 2000). Why would we want to build such machines in general? On the one hand, there is the motive that learning to generate certain intelligent behaviors in artificial systems will help us to understand these behaviors in detail. That is, in our research we devise explanatory models in the form of computer simulations to obtain a better understanding of cognitive and social factors of communication. On the other hand, building conversational machines is expected to help make communication between humans and machines more intuitive.

2.1 Machines as Intentional Systems

Building a machine that can exhibit or simulate rational behavior (*as if* it were an agent acting rationally to further its goals in the light of its beliefs) leads us to look at machines as intentional systems, i.e. systems that perceive changes in the world, represent mental attitudes (like beliefs, goals, etc.), and reason about mental attitudes in order to arrive at decisions on how to plan actions and act.

Research approaches towards modeling mental attitudes and practical reasoning are frequently based on functional models of planning and choosing actions by means-ends analysis, mainly in versions of the belief-desire-intention paradigm (BDI) (Rao and Georgeff 1991).[2] The basic idea is the description of the internal working state of an agent by means of intentional states (beliefs, desires, intentions) as well as the layout of a control architecture that allows the agent to choose rationally a sequence of actions on the basis of their representations. By recursively elaborating a hierarchical plan structure, specific intentions are generated until, eventually, executable actions are obtained (Wooldridge 2002).

Modeling intentional states is based on their symbolic representation. One of its assets is the flexibility it provides for planning and reasoning. In beliefs, for instance, facts concerning the world may be stored that an agent is not (or no longer) able to perceive at the moment, which, however, have effect on the agent's further planning. It is a difference, though, whether an agent draws conclusions simply on the basis of his beliefs and desires or whether he makes use of them – with a corresponding cognitive representation – recognizing them as his own. In many cases such differentiation may not have functional advantages. An agent should be expected, however, to represent intentional states explicitly as being his own ones, if he must also record and deal specifically with other agents' intentional states.

2.2 Bielefeld Max Project

In our research laboratory at Bielefeld taking a cognitive modeling approach scientific enquiry and engineering are closely intertwined. Creating an artificial system that replicates aspects of a natural system can help us understand the internal mechanisms that lead to particular effects. Special for our approach is that we are not just building and studying intelligent functions in separate. Over many years, we have attempted to build coherent comprehensive systems integrating both symbolic and dynamic system paradigms, one of them "Max".

[2] The BDI approach comes from Michael Bratman (Bratman 1987); one of its fundamentals can be traced back to the work of Daniel Dennett (Dennett 1987) on the behavior of intentional systems.

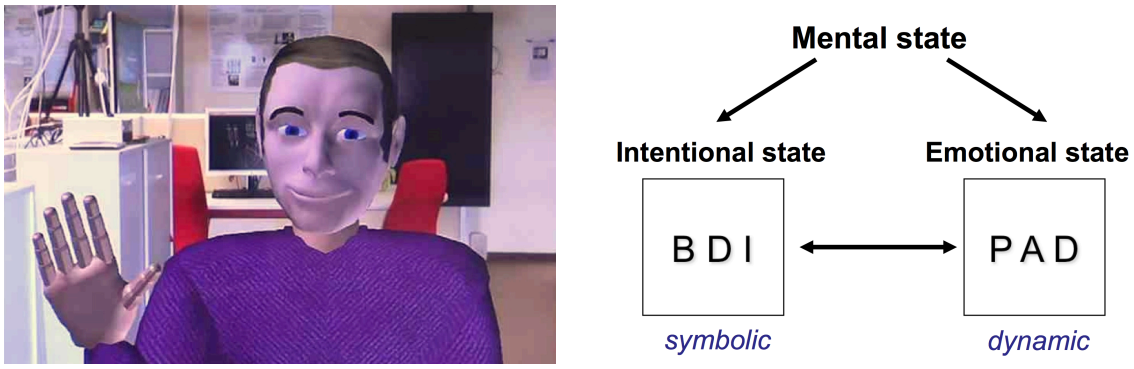

Fig. 1 Virtual human Max: outer appearance and schematic view of internal state

Max is a "virtual human" – an artificial agent embodied in virtual reality with a person-like appearance (see Fig. 1, left). By means of microphones and video cameras or tracker systems, Max can "hear" and "see" his human interlocutors and process verbal instructions and gestures. Max is equipped with verbal conversational abilities and can employ his virtual body to exhibit nonverbal behaviors in face-to-face interaction. With a modulated synthetic voice and an animated face and body, Max is able to speak and gesture, and to mimic emotions (Kopp and Wachsmuth 2004). The face of Max is computer-animated by simulated muscle effects and displays lip-synchronous speech, augmented by eyebrow raises and emotional facial expression. Max's articulated body is driven by a kinematic skeleton (comprising roughly one-third of the degrees of freedom of the human skeleton), with synchronized motion generators giving a realistic impression for his body movements. Emotional expression (which also includes voice modulation) is driven by a dynamic system, which responds to various kinds of stimuli (external: seeing faces, bad words, etc.; internal: goal achievement or failure, resulting in positive resp. negative emotions), and which defines the agent's explicit emotional state over time in pleasure-arousal-dominance (PAD) space (Becker et al. 2004). The agent is controlled by a cognitive architecture (Sect. 2.3) which is based on the symbolic belief-desire-intention (BDI) approach to modeling rational agents, while integrating concurrent reactive behaviors and emotions. Thus the mental state of Max is comprised by an intentional as well as an emotional state (see Fig. 1, right).

With the Bielefeld Max project we investigate the details of face-to-face interaction and how it is possible to describe them – in parts – so precisely that a machine can be made to simulate them. This means that collecting insights about the functioning of human cognitive interaction is an important focus of our work. A technical goal is also the construction of a system as functional and convincing as possible that may be applied in different ways.

2.3 Cognitive Architecture

To organize the complex interplay of sensory, cognitive and actoric abilities, a cognitive architecture has been developed for Max (Leßmann et al. 2006), aiming at making his behavior appear believable, intelligent, and emotional. Here, 'cognitive' refers to the structures and processes underlying mental activities, including attentive processes. Bearing a functional resemblance to the links that exist between perception, action, and cognition in humans, the architecture has been designed for performing multiple activities simultaneously, asynchronously, in multiple modalities, and on different time scales. It provides for reactive and deliberative behaviors running concurrently, with a mediator resolving conflicts in favor of the behavior with the highest utility value.

Figure 2 gives an outline of the cognitive architecture of Max. Explicitly represented goals (desires), which may be introduced through internal processing as well as by external influences, are serving as "inner motivation" triggering behavior. Max can have several desires at the same time, the highest-rated of which is selected by a utility function to become the current intention. The BDI interpreter determines the current intention on the basis of existing beliefs, current desires as well as options for actions.

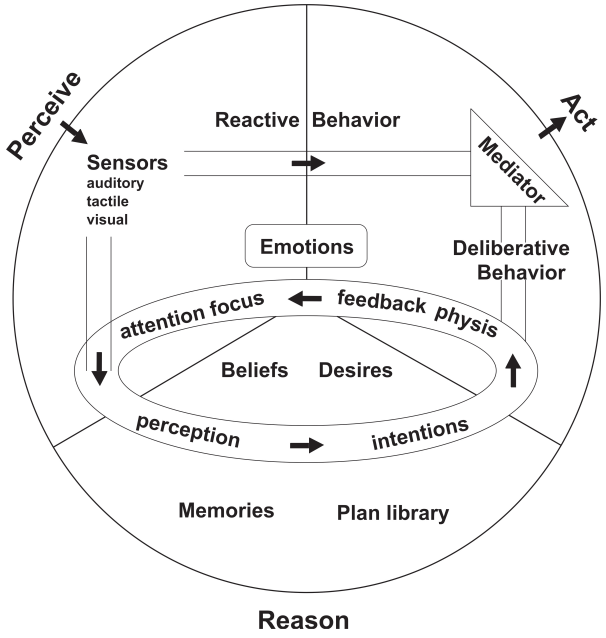

Fig. 2 Outline of the cognitive architecture of Max (Reproduced from Leßmann et al. 2006)

Options for actions are available in a plan library in the form of abstract plans that are described by preconditions, context conditions, consequences that may be accomplished, and a priority function. Plan selection is further influenced by the current emotional state (in PAD space, see above) in that the emotion is used as precondition and context condition of plans to choose among alternative actions. If a concrete plan drawn up on the basis of these facts has been executed successfully, the related goal will become defunct.

The conduct of dialogue is based on an explicit modeling of communicative functions related to, but more specific than, multimodal communicative acts (Poggi and Pelachaud 2000) and generalizing speech act theory (Searle and Vanderveken 1985). A communicative function explicates the functional aspects of a communicative act on three levels (interaction, discourse, content) and includes a performative reflecting the interlocutor's intention, with a dialogue manager controlling reactive behaviors and creating appropriate utterances in response (Kopp et al. 2005). Dialogue is performed in accordance with the mixed initiative-principle, this means, for instance, that in case the human fails to answer, Max himself takes the initiative and acts as the speaker. The plan structure of the BDI system makes it possible to assert new goals during the performance of an intention that may replace the current intention, provided it has a

higher priority. If the previous intention is not specifically abandoned in this process and its context conditions are still valid, it will become active again after the interruption.

2.4 Max as a Museum Guide

Since 2004 Max has been employed as a museum guide in a public computer museum (the Heinz Nixdorf MuseumsForum in Paderborn), taking the step from a research prototype to a system being confronted with many visitors in a rich real-world setting (Kopp et al. 2005). Displayed on a large projection screen (Fig. 3), Max provides the visitors with various information and engages them in a conversation. For instance, greeting a group of visitors, he could say: “Max, that’s me. I’m an artificial person that can speak and gesture. I am artificial, but I can express myself just like you...” A visitor might ask Max “How is the weather?” and Max would then access the current weather forecast in the internet and read it to the visitor. Altogether, this research has embarked on the goal of building embodied agents that can engage with humans in face-to-face conversation and demonstrate many of the same communicative behaviors as exhibited by humans.

Fig. 3 Max interacting with visitors in the Heinz-Nixdorf-MuseumsForum

A screening was done after the first 7 weeks of Max's employment in the computer museum (Kopp et al. 2005). Statistics was based on log files anonymously recorded from dialogues between Max and visitors to the museum. Among other aspects, the data were evaluated with respect to the successful recognition of a communicative function, that is, whether Max could associate a visitor's want with an input. We found that Max recognized a communicative function in nearly two-thirds of all cases. Even when Max had sometimes recognized an incorrect communicative function (as humans may also do), we may conclude that in these cases Max conducted sensible dialogue with the visitors. In the other one-third of cases, Max did not turn speechless but simulated "small talk" by employing commonplace phrases, still tying in visitors with diverse kinds of social interaction.

In some sense, Max could be attributed rational behavior (*as if* he were an agent acting rationally), namely, "minimal rationality" (Dretske 2006): This notion requires not only that behavior be under the causal control of a representational state, but that it be explained by the content of that representational state. Minimal rationality, so to say, requires that what is done is done for *a reason* (not necessarily a good reason). In light of the above statistics on Max's service as a museum guide, Max's answers were given for a reason (in the fulfillment of communicative goals Max associated with visitors' inputs) in many cases. That is, his behavior might be termed "minimally rational" in Dretske's sense (at least in an *as-if* sense – by the way the agent was programmed).

3 From Conversation to Cooperation

The above explained ideas about Max as a conversational machine are relevant also for embodied cooperative systems, i.e. systems acting together with others to accomplish shared goals by employing verbal and nonverbal behaviors in coordinating their actions. Such systems may be embodied as robotic agents (e.g. Breazeal et al. 2004) or (such as Max) as humanoid agents projected in virtual reality. If we want to achieve that Max and a human interlocutor mutually engage and coordinate action in solving a joint problem, a central question is how the processes involved interact and how their interplay can be modeled. For example, inter-agent cooperation relies very much on

common ground, i.e. the mutually shared knowledge of the interlocutors. Nonverbal behaviors such as gaze and gestures are important means of coordinating attention between interlocutors and therefore related to both inferring intentions and coordinating actions. Note that the conduct of dialogue is a form of cooperation, because participants have to coordinate their mental states.

In one of our research settings Max was employed to study cooperative dialogue in a construction task, where Max and a human interlocutor solve the joint problem of constructing a model airplane from a 'Baufix' wooden toykit (Leßmann et al. 2006). In this setting the human interlocutor and Max stand opposite each other at a table (see Fig. 4, left). With the exception of the person shown left, the pictured scene is projected virtual reality. On the table, there are different building bricks: bars with three holes or five holes, several screws with colored caps, a cube with holes on all sides, etc., all items will be assembled in the course of the dialogue. The human sees Max and the entire setting three-dimensionally by means of stereo glasses and hears the synthetic voice of Max due to surround sound from hidden speakers.

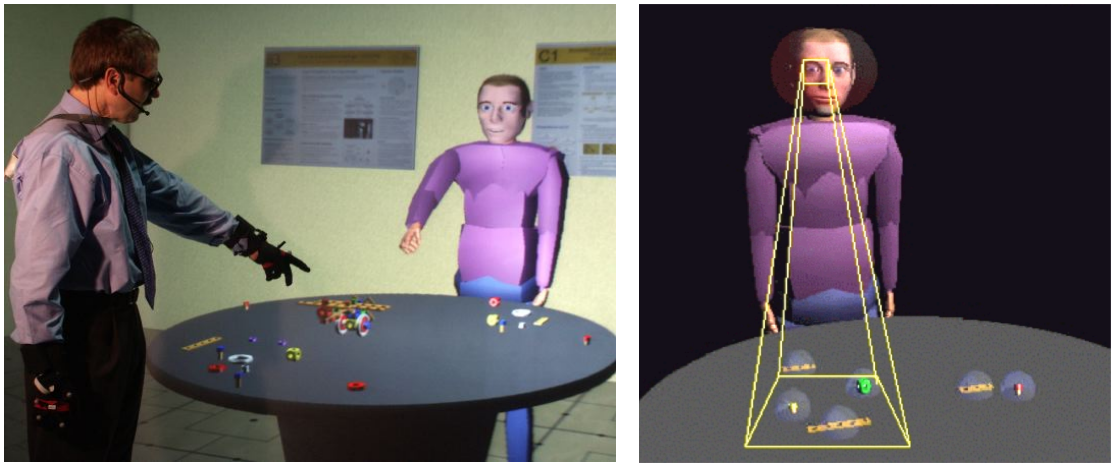

Fig. 4 Cooperative dialogue with Max (left); virtual world perception of Max (right)

To enable the artificial agent to participate in the interaction, Max is equipped with both view sensors for the perception of the virtual environment (see Fig. 4, right) and diverse sensors to capture the signals of the human in the real world. Max 'sees' the human's gaze direction as well as hand and arm movements by means of an infrared tracking system and data gloves, and – via a radio microphone – he 'hears' the human's speaking which he processes with a speech recognizer. The percepts of both worlds

(real and virtual) are integrated to form a coherent perception of the ongoing interaction. Both the human and Max are able to have individual parts assembled by natural language instructions and gestures, carried out in simulation with realistic sounds added. While the human interlocutor in general has the role of proposing the next construction step, Max has knowledge of construction details and helps the other with his role if needed. Thus action roles may switch between the interactants according to their competences. This is done, for example, as follows:[3]

Max Hello, I'm Max. How can I help you? [*smiling*].

Human Explain the assembly of a propeller.

Max Let's do it together. First, a screw has to be inserted into the middle of a three-hole-bar.

Human Insert the yellow screw into the bar [*points to a three-hole-bar, the first screw is inserted into the first hole*].

Max No, that was the wrong hole [*screw slips out*].

Max I'll explain it to you again. A screw has to be inserted into the middle of a three-hole-bar.

Human Insert the screw into the middle of the three-hole-bar [*screw is inserted into the middle of the three-hole-bar chosen before*].

Max Okay, that's correct. Now insert this [*points to the bar with the screw*] into the three-hole-bar.

Human You do it.

Max Okay, I will perform the action [*the objects are connected as instructed*].

Max And now the bars need to be turned crosswise to each other [*shows the rotation using a gesture*], etc.

Fitted to this trial scenario, Max is equipped with limited knowledge of the 'Baufix world' and is capable of planning and reasoning such that he may act as an intelligent assistant. Specialized planners (for constructing Baufix objects) and memories (for storing dynamically updated representations for the state of constructed objects) have been integrated into the cognitive architecture. Further, Max has some

[3] Dialogue translated from German to English.

grammatical rules and a semantic lexicon for processing the meaning of dialogue inputs. Within a limited vocabulary Max is able to talk – including the generation of appropriate gestures –, producing verbal utterances from a repertoire of stereotype statements. These also include the term 'I', without Max having a notion of himself at the current time.[4] Independent of that it could be demonstrated how Max can cope with changing situations that require language, perception, and action to be coordinated so that cooperation between the human and the artificial system takes place with natural efficiency.

4 Coordinating Attention

As was said in the introduction, one of our questions is how joint intentions and cooperation can be modeled and simulated. Attentive processes and sharing attention are important precursors for cooperative interaction in which interactants pursue shared goals by coordinated action plans (joint intentions). Inter-agent cooperation relies much on common ground, one aspect being whether interactants know together that they share a focus of attention (e.g. know that they are both looking at the same target object as illustrated in Figs. 5 and 6 below). This kind of intentionally sharing a focus of attention is referred to as "joint attention" below. It would presuppose interactants to mutually perceive one another and perceive the perceptions of the other, that is, attend to each other. An important means of coordinating attention between interlocutors are nonverbal behaviors such as gaze and gestures. For example, following each other's direction of gaze allows interlocutors to share their attention.

Judged to be crucial for goal-directed behavior, attention has been characterized as an increased awareness of something (Brinck 2003), intentionally directed perception (Tomasello et al. 2005), or as "the temporally-extended process whereby an agent concentrates on some features of the environment to the (relative) exclusion of others" (Kaplan and Hafner 2006). A foundational skill in human social interaction, joint (or shared) attention can be defined as simultaneously allocating attention to a target as a

[4] On how to configure an artificial agent so as to enable him to adopt a first-person perspective see Wachsmuth (2008).

consequence of attending to each other's attentional states, or "re-allocating attention to a target *because* it is the object of another person's attention" (Deák et al. 2001). In contrast to joint perception (the state in which interactants are just perceiving the same target object without further constraints concerning their mental states), the intentional aspect of joint attention has been stressed, in that interlocutors have to deliberately focus on the same target while being mutually "aware" of sharing their focus of attention (Tomasello et al. 2005). If virtual humans are to engage in joint attention with a human interactant, they need to be enabled to meet conditions as described above. For instance, they would need to infer the human interactant's focus of attention from the interactant's overt behaviors. Prerequisites for this are attention detection (e.g. by gaze following) as well as attention manipulation (e.g. by issuing gaze or pointing gestures).

We have investigated joint attention in a cooperative interaction scenario (different from the one described in Sect. 3) with the virtual human Max, where again the human interlocutor meets the agent face-to-face in virtual reality. The human's body and gaze are picked up by infrared cameras and an eye-tracker (mounted on the stereo glasses for three-dimensional viewing), informing Max where the interlocutor is looking at; this way Max can follow the human's gaze (see Fig. 5). For instance, when the human focuses on an object, Max can observe the human's gaze alternating between an object and Max's face and attempt to establish joint attention, by focusing on the same object. Or, initiating a bid for joint attention, Max can choose an object and attempt to draw the attention of his interlocutor to the object by gaze and pointing gestures until joint attention is established.

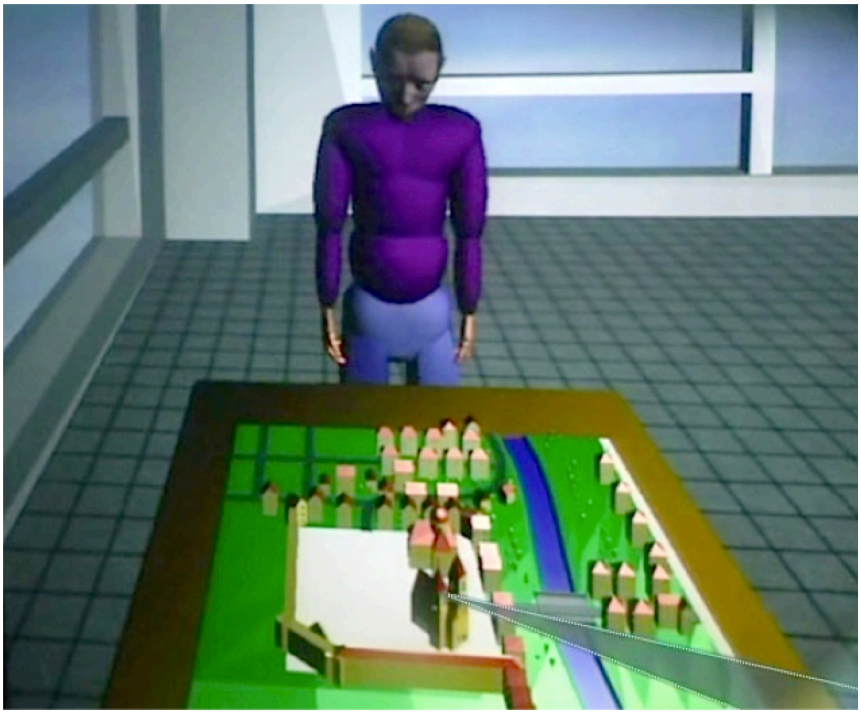

Fig. 5 Max can pick up the human's gaze by means of eye-tracking

So, which inferences exactly need to be drawn to establish joint attention by aligning the mental states of cooperating agents? Pfeiffer-Leßmann and Wachsmuth (2009) describe a formal model which specifies the conditions and cognitive processes that underlie the capacity for joint attention. In accordance with Tomasello et al. (2005) joint attention is conceived of as an intentional process. Our model provides a theoretical framework for cooperative interaction with a virtual human (in our case: Max). It is specified in an extended belief-desire-intention modal logic, which accounts for the temporally-extended process of attention (Kaplan and Hafner 2006) in interaction between two intentional agents, during which agents' beliefs may change. To account for such dynamics of an agent's beliefs in our model, the logic was extended to include activation values. The idea is that activation values influence the beliefs' accessibility for mental operations, resulting in an overall saliency of a belief (and likewise other intentional states). For instance "increased awareness" (Brinck 2003) can be modeled by use of activation values.

In order to account for these ideas, the above described cognitive architecture (Sect. 2.3) adopting the BDI paradigm of rational agents (Rao and Georgeff 1991) has been augmented by incorporating a partner model to account for the agent's perspective

on its interlocutor, as well as a dynamic working memory. The working memory stores the changing beliefs (and other intentional states) of the agent, and also the target objects that may be in the agent's focus of attention. Activation values are used as a measure for saliency, i.e. an object with a higher activation value is more salient than one with a lower activation value. Whenever an object gets in the agent's gaze focus (see Fig. 4, right) or is subject to internal processing, activation values are increased.

To establish joint attention, an agent must employ coordination mechanisms of understanding and directing the intentions underlying the interlocutor's attentional behavior, such as: tracking the attentional behavior of the other by gaze monitoring; deriving candidate objects the interlocutor may be focusing on; inferring whether attentional direction cues of the interlocutor are uttered intentionally; reacting instantly, as simultaneity is crucial in joint attention; and in response employing an adequate overt behavior which can be observed by the interlocutor. Meeting these conditions, Pfeiffer-Leßmann and Wachsmuth (2009) describe the mental state required for an agent i to believe in joint attention while focusing conjointly with its interlocutor j on a certain target ϑ (theta). While we won't go into detail here, see Fig. 6 and brief explanations following for an illustration of these ideas.

Figure 6 illustrates the following (BEL, GOAL, and INTEND are modal connectives in the logic used for modeling beliefs, goals, intentions and attentional states):

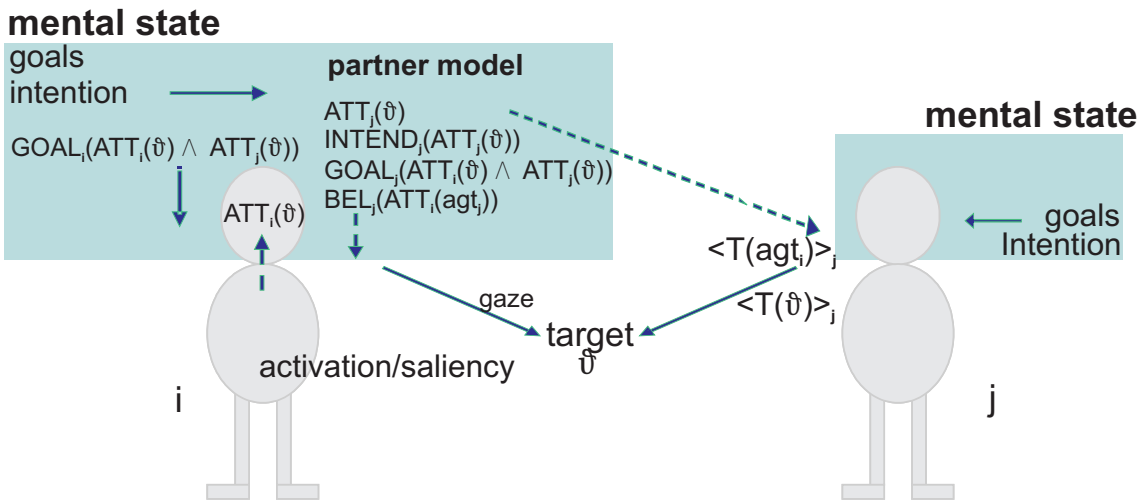

Fig. 6 Joint attention focusing on a target object ϑ from artificial agent i's perspective (Reproduced from Pfeiffer-Leßmann and Wachsmuth 2009)

If agent i (here artificial agent Max) attends (focuses attention: ATT) to a target ϑ and has the goal that both, agents i (Max) and j (human interactant) jointly attend to the same target ϑ, then agent i needs to infer (and assert in the partner model)

– that agent j (also) attends to the target ϑ
– that agent j intends (INTEND) to attend (intentionally attends) to the target ϑ
– that agent j adopts the goal that both agents jointly attend to the target ϑ, and
– that agent j (human interactant) believes (BEL) that (artificial agent) i attends to (human interactant) j

T (test-if) pertains to test-actions that are to infer if human interactant j focuses attention on agent i while simultaneously (observed by gaze-alternation) allocating attention to target ϑ. If agent i (Max) perceives the interlocutor's behavior as a test-action and is able to resolve a candidate target object, the agent infers that the interlocutor's focus of attention resides on that target.[5] The formalization provides a precise means as to which conditions need to be met and which inferences need to be drawn to establish joint attention by aligning the mental states of cooperating agents.

Going from this formal model to the implemented system, some heuristics had to be used, for instance when the human interlocutor focuses several times (or for an extended duration) on an object, the agent interprets this as the attention focus being intentionally drawn upon the target, or that the addressee's response to an agent initiating joint attention needs to take place within a certain time frame. To follow this up, an eye-tracker study was conducted (Pfeiffer-Leßmann et al. 2012) examining dwell times (fixation durations) of referential gaze during the initiation of joint attention, the results of which further contribute to making our formal model of joint attention operational.

[5] For further detail, the formal definition of joint attention, and the specification of epistemic actions that lead to the respective beliefs and goals see Pfeiffer-Leßmann and Wachsmuth (2009).

5 Conclusion: From Tool to Partnership

From the perspective of artificial intelligence, this contribution has addressed embodied cooperative systems, i.e. embodied systems exhibiting cooperative behavior by taking on (some of) the goals of another individual and acting together with the other to accomplish these goals employing verbal and nonverbal communicative behaviors. The questions we set out to address from this perspective were the following: (1) How can joint intentions and cooperation be modeled and simulated? (2) Can we attribute joint intention to a system or team involving both, a human and an artificial agent?

Cooperation, as was said in Sect. 1, involves adopting (some of) the goals of another individual and acting together with the other to achieve these shared goals. Joint intention refers to the ability of interactants to represent coordinated action plans for achieving their common goal in joint activity.

Taking the virtual humanoid agent "Max" as an example of an embodied cooperative system, we introduced first steps towards how joint intentions and cooperation can be modeled and simulated. We described a cognitive architecture, based on the BDI paradigm of rational agents and augmented by a partner model accounting for the agent's perspective on its interlocutor (representing the inferred intentional states in the sense of 'cognitive' ToM), as well as a dynamic working memory storing the changing intentional and attentive states of the agent. This cognitive architecture enables Max to engage in joint activities (conducting conversation, solving a joint problem) with human interlocutors. Further we have outlined which inferences need to be drawn to establish joint attention (the common ground of interactants knowing together that they share a focus of attention) by aligning the mental states of cooperating agents. While a complete model of joint intention remains to be done, we used the case of joint attention to lay out how an artificial agent can represent goals and intentions of his human interactant in a partner model that could be employed for representing coordinated action plans in the plan structure of the BDI system. Thus we could provide some preliminary insights on question (1) as to how joint intentions and cooperation can be modeled and simulated.

Our second question (can we attribute joint intention to a system or team involving both, a human and an artificial agent?) is more complicated since it involves

the idea of partnership, i.e. a relationship between two (or more) individuals working or acting together. There is accumulating evidence that in the context of cooperative settings, the view that humans are users of a certain "tool" has shifted to that of a "partnership" with artificial agents, insofar they can in some sense be considered as being able to take initiative as autonomous entities (Breazeal et al. 2004; Negrotti 2005). According to Negrotti (2005), a true partner relationship is to be conceived as a peer-based interaction, wherein each partner can start some action unpredicted by the other. Looking at Max, we note that the cooperative interaction in the above described scenarios is characterized by a high degree of interactivity, with frequent switches of initiative and roles of the interactants. In consequence, though being goal-directed, the interaction with Max appears fairly unpredictable. Thus Max appears to be more than a tool (thing) entirely at our disposal and under our control.

So, in light of what has been said above, can we attribute joint intention to a system or team involving both, a human and an artificial agent?

Perhaps one day. There is a long way to go, though. We have to acknowledge that state-of-art agent technology is still far from being sufficiently sophisticated to implement all the behaviors necessary for a cooperative functionality and in particular joint intention in a coherent technical system.[6] But it has to be realized that artificial systems may increasingly take on functions which were reserved to human beings so far and thus seem to become more human-like. For instance, in recent work (Mattar and Wachsmuth 2014), a person memory was developed for Max which allows the agent to use personal information remembered about his interlocutors from previous encounters which, as evaluations have shown, makes him a better conversational partner in the eyes of his human interlocutors.

Also to be noted, the desires of Max do not originate in "real needs" that Max might have; they were programmed functionally equivalent to intentional states we would attribute to a real person, resulting in behaviors that appear somewhat rational. Another programmed "need" of Max is that he demands his conversational partners to be polite. The emotional state of Max (see Sect. 2) is negatively influenced by inputs containing ungracious or politically incorrect wordings ("no-words") which, when

[6] Note that, even when we have attempted to build a coherent comprehensive system, not all aspects described in this article have been integrated in one system, that is, different versions of "Max" were used to explore the above ideas in implemented systems.

repeated, can eventually trigger a plan causing the agent to leave the display and stay away until the emotion has returned to a balanced state (an effect introduced to de-escalate rude visitor behavior in the museum). The period of absence can either be shortened by complimenting Max or extended by insulting him again, see (Becker et al. 2004). Altogether, this kind of behavior of Max may be taken as beginnings of moral judgement.

In conclusion: If we want to construct artificial systems that are helpful to humans and interact with us like "partners", then such systems should be able to understand and respond to the human's wants – infer and share our intentions – in order to be assistive in a given situation. It may be asked if it makes a big difference for embodied cooperative systems to be helpful whether their understandings and intentions (and the intentions they share with us) are real or "*as-if*" (Stephan et al. 2008).

Acknowledgments The research reported here draws on the contributions by the members of Bielefeld University's artificial intelligence group which is hereby gratefully acknowledged. Thanks also to Catrin Misselhorn and an anonymous referee for helpful comments to improve the text. Over many years this research has been supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Centers 360 (Situated Artificial Communicators) and 673 (Alignment in Communication), the Excellence Cluster 277 CITEC (Cognitive Interaction Technology), and the Heinz Nixdorf MuseumsForum (HNF). This paper is a preprint version of an article published by Springer. The original publication is available at: http://link.springer.com/chapter/10.1007%2F978-3-319-15515-9_4

References

Becker, Christian, Stefan Kopp, and Ipke Wachsmuth. 2004. Simulating the emotion dynamics of a multimodal conversational agent. In *Affective dialogue systems*, ed. E. André, L. Dybkjaer, W. Minker, and P. Heisterkamp, 154-165. Berlin: Springer.

Boukricha, Hana, Nhung Nguyen, and Ipke Wachsmuth. 2011. Sharing emotions and space – empathy as a basis for cooperative spatial interaction. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents* (IVA 2011), ed. S. Kopp, S. Marsella, K. Thorisson, and H.H. Vilhjalmsson, 350–362. Berlin: Springer.

Bratman, Michael E. 1987. *Intention, plans, and practical reason*. Harvard: Harvard University Press.

Bratman, Micheal E. 1992. Shared cooperative activity. *Philosophical Review* 101(2): 327–341.

Breazeal, Cynthia, Andrew Brooks, David Chilongo, Jesse Gray, Guy Hoffman, Cory Kidd, Hans Lee, Jeff Lieberman, and Andrea Lockerd. 2004. Working collaboratively with humanoid robots. In *Proceedings of Humanoids 2004*, Los Angeles.

Brinck, Ingar. 2003. The objects of attention. In *Proceedings of ESPP 2003* (European Society of Philosophy and Psychology, Torino, Italy 9–12 July 2003), 1–4.

Cassell, Justine, J. Sullivan, S. Prevost, and E. Churchill (eds.). 2000. *Embodied conversational agents*. Cambridge, MA: MIT Press.

Deák, Gedeon O., Ian Fasel, and Javier Movellan. 2001. The emergence of shared attention: Using robots to test developmental theories. In *Proceedings of the first international workshop on epigenetic robotics*, Lund University Cognitive Studies, vol. 85, 95–104.

Dennett, Daniel C. 1987. *The intentional stance*. Cambridge, MA: MIT Press.

Dretske, Fred I. 2006. Minimal rationality. In *Rational animals?*, ed. Susan L. Hurley, and Matthew Nudds, 107-116. Oxford: Oxford University Press.

Kaplan, Frédéric, and Verena V. Hafner. 2006. The challenges of joint attention. *Interaction Studies* 7(2): 135–169.

Kopp, Stefan, and Ipke Wachsmuth. 2004. Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds* 15: 39-52.

Kopp, Stefan, Lars Gesellensetter, Nicole C. Krämer, and Ipke Wachsmuth. 2005. A conversational agent as museum guide – design and evaluation of a real-world application. In *Intelligent virtual agents*, ed. Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier, and Thomas Rist, 329-343.

Berlin: Springer.

Krämer, Nicole C. 2008. Theory of Mind as a theoretical prerequisite to model communication with virtual humans. In *Modeling communication with robots and virtual humans*, ed. Ipke Wachsmuth and Günther Knoblich, 222-240. Berlin: Springer.

Leßmann, Nadine, Stefan Kopp, and Ipke Wachsmuth. 2006. Situated interaction with a virtual human – perception, action, and cognition. In *Situated communication*, ed. Gert Rickheit and Ipke Wachsmuth, 287-323. Berlin: Mouton de Gruyter.

Mattar, Nikita, and Ipke Wachsmuth. 2014. Let's get personal: Assessing the impact of personal information in human-agent conversations. In *Human-computer interaction*, ed. M. Kurosu, 450–461. Berlin: Springer.

Negrotti, Massimo. 2005. Humans and naturoids: from use to partnerships. In *Yearbook of the artificial Vol. 3, cultural dimensions of the user*, ed. Massimo Negrotti, 9-15. Bern: Peter Lang European Academic Publishers.

Pfeiffer-Leßmann, Nadine, and Ipke Wachsmuth. 2009. Formalizing joint attention in cooperative interaction with a virtual human. In *KI 2009: Advances in artificial intelligence*, ed. B. Mertsching, M. Hund, and Z. Aziz, 540–547. Berlin: Springer.

Pfeiffer-Leßmann, Nadine, Thies Pfeiffer, and Ipke Wachsmuth. 2012. An operational model of joint attention – timing of gaze patterns in interactions between humans and a virtual human. In *Proceedings of the 34th annual conference of the Cognitive Science Society*, ed. N. Miyake, D. Peebles, and R.P. Cooper, 851–856. Austin, TX: Cognitive Science Society.

Poggi, Isabella, and Catherine Pelachaud. 2000. Performative facial expression in animated faces. In *Embodied conversational agents*, ed. J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, 155–188. Cambridge, MA: MIT Press.

Premack, David, and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 4: 512-526.

Rao, A.S., and M.P. Georgeff. 1991. Modeling rational agents within a BDI-architecture. In *Principles of knowledge representation and reasoning*, ed. J. Allen, R. Fikes, and E. Sandewall, 473–484. San Mateo CA: Morgan Kaufmann.

Schank, Roger C. 1971. Finding the conceptual content and intention in an utterance in natural language conversation. In *Proceedings of IJCAI 1971* (International Joint

Conference on Artificial Intelligence, London, England 1–3 Sept 1971), 444-454.

Searle, John R., and Daniel Vanderveken. 1985. *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.

Stephan, Achim, Manuela Lenzen, Josep Call, and Matthias Uhl. 2008. Communication and cooperation in living beings and artificial agents. In *Embodied communication in humans and machines*, ed. Ipke Wachsmuth, Manuela Lenzen, and Günther Knoblich, 179–200. Oxford: Oxford University Press.

Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28: 675-691.

Wachsmuth, Ipke. 2008. 'I, Max' – Communicating with an artificial agent. In *Modeling communication with robots and virtual humans*, ed. Ipke Wachsmuth and Günther Knoblich, 279-295. Berlin: Springer.

Wooldridge, Michael. 2002. *An introduction to multiagent systems*. Chichester: Wiley.