Chapter 15
Group Cognition and Collaborative AI

Janin Koch and Antti Oulasvirta

Abstract Significant advances in artificial intelligence suggest that we will be using intelligent agents on a regular basis in the near future. This chapter discusses group cognition as a principle for designing collaborative AI. Group cognition is the ability to relate to other group members' decisions, abilities, and beliefs. It thereby allows participants to adapt their understanding and actions to reach common objectives. Hence, it underpins collaboration. We review two concepts in the context of group cognition that could inform the development of AI and automation in pursuit of natural collaboration with humans: conversational grounding and theory of mind. These concepts are somewhat different from those already discussed in AI research. We outline some new implications for collaborative AI, aimed at extending skills and solution spaces and at improving joint cognitive and creative capacity.

15.1 Introduction

The word 'collaboration' is derived from the Latin *col-* ('together') and *laborare* ('to work'). The idea of a machine that collaborates with humans has fired the imagination of computer scientists and engineers for decades. Already J.R. Licklider wrote about machines and humans operating on equal footing and being able to 'perform intellectual operations much more effectively than a man alone' [60].

If there is a shared tenet among the visionaries, it is that the more complex the activities become – consider, for example, planning, decision-making, idea generation, creativity, or problem-solving – the more beneficial collaboration is. However,

J. Koch (✉) · A. Oulasvirta
Department of Communications and Networking, School of Electrical Engineering, Aalto University, Espoo, Finland
e-mail: janin.koch@aalto.fi

A. Oulasvirta
e-mail: antti.oulasvirta@aalto.fi

© Springer International Publishing AG, part of Springer Nature 2018

J. Zhou and F. Chen (eds.), *Human and Machine Learning*, Human–Computer Interaction Series, https://doi.org/10.1007/978-3-319-90403-0_15

janin.koch@aalto.fi

although collaboration has received attention from research on automation, robotics, Artificial Intelligence (AI), and Human-Computer Interaction (HCI), it can be safely said that most technology is not yet collaborative in the strong sense of the term. Humans are mainly in a commanding role or probed for feedback, rather than parties to a mutually beneficial partnership. There is much that could be done to better use human abilities in computational processes, and vice versa.

The topic of this chapter is group cognition: the ability to bring about a common understanding among agents; relate to other agents' decisions, abilities, and beliefs; and adapt one's own understanding toward a common objective [82]. This goes beyond the common notion of a computer 'understanding' human intents and actions, and highlights the necessity of contextual awareness, the ability of communicating reasoning behind actions to enable valuable contributions [51]. This, we argue, would result in human–machine collaboration that not only is more efficient but also is more equal and trustworthy.

We find group cognition particularly promising for re-envisioning what AI might need to achieve for collaboration, because it meshes with a strong sense of the concept of collaboration. Group cognition emerges in interaction when the group members involved, humans or computers, share knowledge and objectives and also dynamically and progressively update their understanding for better joint performance. This captures one aspect of the essence of machines that can be called collaborative.

Group cognition points towards various abilities necessary for collaboration. In this chapter we ask which of these abilities are needed for collaborative AI's. Among the many fields one might consider in the context of collaborative behaviour, management psychology presents an extensive body of research on how team members collaborate to solve common problems together [46], while developmental psychology has looked more closely at collaboration as an evolving behaviour in humans [32]. By comparison, AI and HCI research has looked at collaboration from the principal–agent perspective [65], in terms of dialogue and initiative [5], and as computer-mediated human–human collaboration [35]. Perhaps the most significant advances related to algorithmic principles of collaboration in the field of computer science have been made in the field of interactive intelligent systems [9, 81] and human–robot interaction [77]. However, on account of its roots in psychology and education, the concept of group cognition is rarely referred to within computational and engineering sciences.

To this end, we provide definitions, examples, and discussion of implications of the design of such an AI, where 'AI' refers mainly to machine learning-based intelligent systems though not being limited to that sense. We further discuss two key aspects of group cognition, by borrowing the concepts of conversational grounding and theory of mind. Even though these concepts overlap somewhat with each other, their use in combination does not map onto any existing concept in AI research.

Recent advances in AI have shown capabilities that are clearly relevant for group cognition, such as intent recognition [59], human-level performance in problem-solving [23], and cognitive artificial intelligences [90]. However, these capabilities do not trivially 'add up to' a capability of group cognition. In contrast to previous thought, wherein machines have often been described as extended minds or

janin.koch@aalto.fi

'assistants', we hold that a system capable of group cognition would better understand human actions as part of a joint effort, align its actions and interpretations with the interpretation of the group, and update them as the activity evolves. A sense of dependability and common cause would emerge, which would improve the trustworthiness of such collaboration. In this way, a system capable of group cognition could participate in more open-ended, or ill-defined, activities than currently possible.

15.2 Definitions of Collaboration

We start by charting some definitions of collaboration. This groundwork serves as a basis for reflecting on group cognition as a theory of social behaviour. In social sciences and philosophy, the key phenomenon in collaboration is called *intersubjectivity*. Intersubjectivity refers to how two or more minds interrelate: understand each other and work together from their individual cognitive positions [83]. Some well-known social theories related to intersubjectivity are *mediated cognition* [87], *situated learning* [56], *knowledge building* [45], and *distributed cognition* [42]; D.J. Wood and B. Gray present an overview of differences among these perspectives [91]. We illustrate these differences with reference to a small selection of commonly accepted definitions.

Collaborative work has been defined within the domain of organisational work as 'a mutually beneficial relationship between two or more parties who work toward common goals by sharing responsibility, authority, and accountability for achieving results' [18]. This definition is used to understand collaboration in companies and other organisations, and the focus has been mainly on the outcome and values of team collaboration. Knowledge discovery in problem-solving is emphasised in the definition of collaboration as 'a continued and conjoined effort towards elaborating a "joint problem space" of shared representations of the problem to be solved' [7]. A third definition we wish to highlight focuses on differences among the contributing actors. Here, collaboration is 'a process through which parties who see different aspects of a problem can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible' [38].

In this chapter, we build on a fourth definition, from Roschelle et al., who define collaboration as 'a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem' [72]. This definition builds on the notion of collaboration as a cognitive action but also includes aspects of the previously mentioned definitions. The latter definition originated in the field of collaborative learning. Some empirical evidence exists that such collaborative learning enhances the cognitive capabilities of the people involved, allowing them as a team to reach a level of cognitive performance that exceeds the sum of the individuals' [7].

Collaboration also has to be distinguished from co-operation, a notion that is at times used to characterise intelligent agents. Roschelle et al. suggest that co-operative work is 'accomplished by the division of labour among participants, as an

janin.koch@aalto.fi

activity where each person is responsible for a portion of the problem-solving' [72], whereas collaborative learning involves the 'mutual engagement of participants in a coordinated effort to solve the problem together' [72]. Co-operation and collaboration differ also in respect of the knowledge involved and the distribution of labour. To co-operate means at least to share a common goal, towards whose achievement each participant in the group will strive. But this is compatible with dividing the task into subtasks and assigning a specific individual (or subgroup) responsibility for completing each of these. We can conclude, then, that 'to collaborate' has a stronger meaning than 'to co-operate' (in the sense of pursuing a goal that is assumed to be shared). The former involves working together in a more or less synchronous way, in order to gain a shared understanding of the task. In this sense, co-operation is a more general concept and phenomenon than collaboration.

Collaboration is a specific form of co-operation: co-operation works on the level of tasks and actions, while collaboration operates on the plane of ideas, understanding, and representations. In light of these definitions, research on group cognition can be viewed as an attempt to identify a necessary mechanism behind humans' ability to collaborate.

15.3 Group Cognition: A Unifying View of Collaboration

The core research goal on group cognition has been to shed light on cognitive abilities and social phenomena that together enable what is called 'collaboration'. The widely cited definition of group cognition alluded to above points out three qualities: (1) an ability to converge to a common understanding among agents; (2) an ability to relate to other agents' decisions, abilities, and beliefs; and (3) an ability to adapt one's own understanding toward a common objective during collaboration [82].

Research on group cognition has focused mostly on learning and ideation tasks in small groups (of people). A group's shared knowledge is claimed to be constructed through a process of negotiating and interrelating diverse views of members. Participants learn from each other's perspectives and knowledge only by accepting the legitimate role of each within the collaboration. This distinguishes group cognition from concepts such as extended cognition [36], wherein other participants are vehicles for improving or augmenting the individual's cognition rather than legitimate partners. The implication for AI is that, while a system for extended cognition would allow a person to complete work more efficiently by lessening the cognitive load or augmenting cognitive abilities, a 'group-cognitive system' would complement a human partner and take initiative by constructing its own solutions, negotiating, and learning with and for the person. It would not only improve the cognitive abilities of the human but enhance the overall quality of joint outcomes.

In group cognition, participants construct not only their own interpretations but interpretations of other participants' beliefs [82]. This distinguishes group cognition from previous concepts of collaboration such as conversational grounding [20]. Group cognition is not so much the aggregation of single cognitions as the outcome

janin.koch@aalto.fi

of synchronisation and coordination of cognitive abilities among the participants, cohering via interpretations of each other's meanings [86]. It has been argued that groups that achieve this level feel more ownership of the joint activity [24, 63]. This observation has encouraged studies of group cognition in childhood development, work, and learning contexts [3, 13].

In contrast to isolated concepts traditionally used in HCI and AI today, group cognition may offer a theoretical and practical framing of cognitive processes underpinning human-with-human collaboration. For machines to become collaborative participants, their abilities must be extended toward the requirements following from attributes of group cognition. This would allow machines to expand their role from the current one of cognitive tools toward that of actual collaborating agents, enabling the construction of knowledge and solutions that go beyond the cognition of each individual participant.

In this chapter, though, we consider mainly dyadic collaboration, involving a human–machine pair. Even though this restricts our scope to a subset of the phenomena encompassed by group cognition, larger groups require additional co-ordination, which is not addressed within the constraints of this chapter.

Taking the definition of group cognition as a foundation for our analysis, we can identify two main aspects of successful human–machine collaboration: (1) the ability of recurrently constructing mutual understanding and meaning of the common goal and interaction context and (2) the ability to interpret not only one's own reasoning but also the reasoning of other participants. In order to discuss these requirements in more detail, we make use of recognised theories from cognitive science and collaborative learning – namely, *conversational grounding* and *theory of mind*. Both theories contribute to a comprehensive view of group cognition. Though the two have considerable overlap, both are necessary if we are to cover the fundamental aspects of group cognition [7, 82].

In the following discussion, we briefly introduce these theories and explain their relation to group cognition. Proceeding from this knowledge, we then present key requirements and explain their potential resolution. Then, in Sect. 15.6, we present current realisations of systems addressing these requirements, identify limitations, and present ideas for future research.

15.4 Conversational Grounding

'Grounding' refers to the ability to create a shared base of knowledge, beliefs, or assumptions surrounding a goal striven toward [8]. Whilst taking grounding to be a complete explanation of collaborative behaviour has been questioned, the concept's explanatory power for constructing meaning in small-scale, dyadic collaboration has been demonstrated in several studies [82].

The term is used in the sense employed by Clark et al. within the tradition of conversational analysis [20]. They argue that common ground and its establishment are the basis for collaboration, communication, and other kinds of joint activity.

janin.koch@aalto.fi

Especially within dyadic interactions, it has informed various theoretical frameworks, even in AI. Among the most prominent are the collaborative model [19], the Mixed-Initiative (MI) framework [5], and theories of collaborative learning [8]. Grounding highlights the necessity for efficient communication to ground the collective understanding by ensuring not only clear expression of the contributions to the collaboration but also correct reception by the addressees. It is thus a basic constitutive activity in human–human communication and collaboration.

It has been claimed that two factors influence success in grounding: purpose and medium [20]. *Purpose* refers to the objective, desire, or emotion that should be conveyed within a collaborative undertaking. The *medium* is the technique to express the current purpose, which includes the cost its application requires. Clark et al. introduce the concept of the 'least collaborative effort' [20], according to which participants often try to communicate as little as possible – but as much as necessary – with the most effective medium to allow correct reception. From this perspective, work on mixed-initiative interaction has addressed mainly the co-ordination of communication, *when* to communicate. Grounding could inform MI and other AI frameworks with regard to how reciprocal understanding among participants could be achieved. To this end, we can identify four key requirements:

(1) Expressing one's own objectives: Grounding is based on successful expression of one's objectives, requirements, and intents that define the purpose of the conversation in the collaborative activity [20]. Achieving this with a computer is not trivial. In a manner depending on the medium, a system has to divide information into sub-elements, which can then be presented to other group members (e.g., a concept into sufficiently descriptive words). Among examples that already exist are dialogue systems applying Belief–Desire–Intention models [48] and theoretical models for purposeful generation of speech acts [21] to construct meaningful expressions of objectives. Also, there is a growing body of research exploring the potentials of concept learning [25, 53], which would enable a machine to combine objectives and communicate or associate them with existing concepts.

(2) Selecting the most effective medium: To collaborate, a participant has to select the medium that can best convey the purposes of the conversation. In human-to-human conversation, a purpose can be expressed in various ways, including verbal and non-verbal communication. The choice of medium depends on the availability of tools, the effort it requires to use the medium, and the predicted ability of the collaborator to perceive the purpose correctly. Tools in this context are all of the means that help to convey the purpose – e.g., speech, pointing, body language, and extended media such as writing or drawing. The effort of using a medium depends on skills and the ability to apply them. In the case of drawing, the effort would include getting a pencil and paper as well as having the ability to draw the intended purpose. Finally, the selection of the medium depends also on the ability of other participants to perceive it correctly. This is related to the ability to physically perceive the medium (for example, hand gestures' unavailability during a phone call) and to the predicted ability to understand the medium. The ability of an intelligent system to select a medium is obviously limited by its physical requirements. While embodied robots share the same space and the same media as a human and can engage in pointing,

janin.koch@aalto.fi

eye movement, or use of voice [62], virtual agents possess limitations in addressing physical elements when referring. On the other hand, virtual agents' enhanced skills with visual, animated, or written representations of information can be exploited as a comparatively strong expressive medium.

(3) Evaluating the effort of an action: H.H. Clark and D. Wilkes-Gibbs introduce the principle of least collaborative effort as a trade-off between the initial effort of making oneself understood and the effort of rephrasing the initial expression upon negative acknowledgement, as in the case of misunderstanding [19]. Previous work on least effort has examined short and precise communication efforts, which favour unique wording as optimal strategy. In contrast, Clark and Wilkes-Gibbs show that least *collaborative* effort does not necessarily follow the same pattern. On account of the joint effort of understanding within collaboration, the interpretation of least effort can be relaxed and groups can also accept wordings with internal references that are not necessarily unique to the given context. This presents both an opportunity and a challenge for machines. The conversation structure of most conversational agents, such as Siri [47], follows the least effort principle, by providing short and specific answers. Extending this to a least-collaborative-effort strategy would imply the ability to connect knowledge with previous and general expressions. N. Mavridis presents 'situated language' to overcome these issues and enable a machine to extend its communication ability to time- and place-dependent references [62].

(4) Confirming the reception of the initial objective: For successful conversational grounding, the group member expressing knowledge not only must find the right medium and dimension for expression but also has to verify correct reception by other members through evidence [20]. This allows the group to create mutual understanding within the process. Evidence for understanding may be positive, indicating that the receiving participant understood, or negative. People often strive for positive evidence of correct reception of their expression, which can be provided either through positive acknowledgement, such as nodding or 'mmm-hmm', or via a relevant next-turn response, which may be an action or expression building on the previous turn(s). Naturally, the reaction to the expression might differ with the medium used. Enabling a machine to evaluate understanding by other group members, therefore, entails new research into not just natural-language processing in relation to natural interaction [11] but also handling of non-verbal behaviour [29].

While grounding refers to the ability to communicate and receive knowledge to find 'common ground', group cognition goes beyond that. It additionally requires reciprocal interpretation of thoughts and intentions, for relation to other group members' decisions and beliefs [7]. In order to highlight this, we borrow from theory of mind as a basis for our analysis in the next part of the chapter.

15.5 Theory of Mind

The ability of interpreting one's own knowledge and understanding as well as interpreting other collaborators' understanding is crucial for successful collaboration in group cognition [31]. Theory of mind is a topic originating from research on

janin.koch@aalto.fi

cognitive development. It focuses on the ability to attribute mental states to oneself and others and to recognise that these mental states may differ [15]. Mental states may be beliefs, intentions, knowledge, desires, emotions, or perspectives, and understanding of these builds the basis for grounding. The ability to interpret others' mental states allows humans to predict the subsequent behaviour of their collaborators, and it thereby enables inferring the others' aims and understandings.

While most research on theory of mind has focused on developmental psychology, a growing body of literature backs up its importance for group behaviour [2] and group cognition [83], suggesting the importance of the concept for human–machine collaboration [15]. Human–machine interaction nevertheless is often limited by the level of ability to interpret the 'mind' of machines, on account of their different, sometimes unexpected, behaviour. People still approach new encounters with technology similarly to approaching other human beings, and attribute their own preconceptions and social structures to them [15, 34]. For reason of machines' inability to interpret their own mental state and that of others, prediction of the behaviour of humans in line with preconceptions often fails.

Three abilities stand out as vital for the development of collaborative AI in this context:

(1) Interpreting one's own mental states: Enabling an intelligent machine to interpret its own mental states requires a computational notion of and access to intentions, desires, beliefs, knowledge, and perspectives. At any point during collaboration, a mental state with regard to another group member may depend on the content of the discourse, the situation, and the information about the current objective.

Most AI applications have been limited to specific tasks, to reduce the complexity of the solution space by decreasing the number of objectives, requirements, and intents involved. However, this also reduces the machine's ability to adapt to changing contexts as found in a discourse, wherein it is necessary to extend the predefined belief space. Recent approaches in collaborative machine learning have constituted attempts to overcome the limitation of single-purpose systems [55]. These allow various information sources, such as sensors, to be integrated into a larger system, for broader knowledge. However, these sources have to be well integrated with each other if they are to create coherent knowledge about a situation [36].

(2) Interpreting others' mental states: Humans constantly strive to attribute mental states to other collaboration participants, to enable prediction of the others' subsequent reactions and behaviours [15]. Such reasoning enables conversations to be incremental. Incremental conversation refers to the ability to follow a common chain of thoughts and forms the basis of any argumentation and subsequent discussion (as in brainstorming). A large body of work on machine learning and AI is related to identifying and predicting human intention [28, 66] and actions [29, 88, 89]. However, this requirement is reciprocal and implies the same needs related to human understanding of the AI mind.

(3) Predicting subsequent behaviour: Similarly to interpretation of another's mental state, prediction of later behaviour can be considered from two sides: Humans apply a certain set of underlying preconceptions to interactions with intelligent

janin.koch@aalto.fi

systems, which often leads to disrupted experiences that arise from unexpected behaviour of the system [15, 77]. Scholars are attempting to identify the information needed for predicting behaviour of machines. In A. Chandrasekaran et al.'s study of human perception of AI minds [15], humans were not able to predict the subsequent behaviour of the AI even when information about the inner mental states, like certainty and attention, of the machine was presented. In contrast, research into machines' prediction of human behaviour has a long history and has already yielded promising results [67, 73].

Group cognition is an interactive process among group members and requires participants to reason about decisions and actions taken by others in order to find common, agreeable ground for progress in the collaboration. While theory of mind explains the former underlying cognitive principles well, it does not explain how this common ground is built. For this reason, we have combined the two theories for our discussion, to elaborate a more comprehensive list of abilities necessary for AIs' engagement in collaboration.

15.6 Challenges for Collaborative AI

The group cognition angle may pinpoint mechanisms necessary for collaborative interaction between humans and artificial agents. In this context, we have highlighted two key concepts – conversational grounding and theory of mind. In summary, group cognition requires both the ability to internalise and constantly update knowledge in line with one's interpretation, as described in theory of mind, and a mutual understanding of the collaboration's purpose, provided through grounding. In the following discussion, we reflect on how these two concepts tie in with current research on AI, highlighting which capabilities may already be achievable by means of existing methods and which still stand out as challenges for future research.

15.6.1 Identify One's Own Mental States

Human–human collaboration is based on the assumption that participants are able to identify their own objectives, knowledge, and intents – in other words, their mental states. Extrapolating intentions from one's own knowledge based on the collaboration interaction and the mutual understanding of the goal is crucial.

Two limitations stand out. Firstly, although there is increasing interest in self-aware AI, most work on the inference of mental states has considered inference of people's mental states while ignoring the necessity of interpreting the machines' 'mental states' [15]. Secondly, because 'common sense' is still out of reach for AI, most (interactive) machine learning and AI systems address only narrow-scoped tasks. This limits their ability to form a complete picture of the situation, inferring and constructing human-relatable intents.

janin.koch@aalto.fi

15.6.2 Select the Communication Medium and Express Objectives

If it is to express objectives and intents, a machine has to select the most efficient way to express them, as suggested in our discussion of grounding. There is a trade-off between the effort it takes for the machine to use a certain communication medium and the chances of the communication being received incorrectly.

The medium of choice for most interactive virtual agents is text. Examples include interactive health interfaces [33, 71] and industrial workflows [39], along with dialogue systems such as chat bots [41] and virtual personal assistants [57]. In recent virtual assistants, text is often transformed into spoken expression. However, the systems usually apply stimulus–response or stimulus–state–response paradigms, which does not suffice for natural speech planning or dialogue generation [62]. Another medium is visual representation via, for example, drawing, sketching, and/or presenting related images. Even if it requires further effort to translate the objectives of a conversation to visual representation, people are especially good at understanding drawings of concepts, even when these are abstract [19]. While virtual agents are starting to use graphics such as emoji or more complex images to convey emotions [30], communication through visual representations, overall, represents an underresearched opportunity in human–machine collaboration. The field of human–robot interaction, meanwhile, has looked at more natural conversational media for expressing objectives or intents [62]. Here, verbal communication is combined with nonverbal communication, such as gaze-based interaction [93], nodding [79], pointing [74], and facial gestures.

However, more studies are needed before we will be able to exploit the potential of gestural and gaze behaviour, along with more graphical representations at different abstraction levels. That work could result in a more efficient medium for communication to humans than is observable in human interaction today.

15.6.3 Confirm the Reception and Interpretation of Objectives

Communication, according to the grounding theory, is successful when a mutual understanding is created. This requires successful reception of the objective expressed. Reception – and acknowledgement of it to the communication partner – is necessary for understanding of mental states and objectives within a group. We can borrow the principle of evidence for reception [20] to state that machines should expect and work with the notion of positive or negative evidence.

Here, negative and positive evidence have a more specific meaning than in the sense of negative and positive feedback familiar from machine learning. Clark et al. identify two possible ways of giving positive evidence, next-turn responses and positive acknowledgement [20]. Next-turn responses are evaluated by looking at the

janin.koch@aalto.fi

coherence between one's own inference and the group member's next-turn responses as well as the initial objectives of the conversation. A.C. Graesser et al., for example, present an intelligent tutoring system that emphasises the importance of next-turn response that is based on learning expectations instead of predefined questions [37]. When interacting with a student, it 'monitors different levels of discourse structure and functions of dialogue moves' [37]. After every answer, it compares the response with the objectives by applying latent semantic analysis, then chooses its communication strategy accordingly. This allows the system to reformulate the question or objective when it perceives that the response does not match expectations. Further examples of such systems are presented by B.P. Woolf [92].

In open-ended tasks such as brainstorming, however, the next-turn response might not have to do with the initial objective so much as with extension or rejection of the intent behind it. In such contexts, humans often fall back to positive and negative acknowledgements. Recognising social signals such as nodding, gaze, or back-channel words of the 'uh-huh' type as positive acknowledgement plays an important role in human interaction and hence is an important ability for a fully collaborative machine. Within the field of human–robot interaction, recognising these signals has been an active research topic for some time [69, 94]. D. Lala et al. have presented a social signal recognition system based on hierarchical Bayesian models that consider nodding, gaze, laughing, and back-channelling as social signals for engagements and acknowledgement [54] with promising results. This approach allows determining which social cues are relevant on the basis of judgements of multiple third-party observers and includes the latent character of an observer as a simulation of personality. The detection of social signals, acknowledgements, would allow a machine to adapt its behaviour and reactions to the other group members.

15.6.4 Interpret the Reasoning of Others

If they are to contribute efficiently to a collaborative effort, group members have to understand the reasoning of the other participants. We use 'reasoning' to mean not merely mental states but also the logic and heuristics a partner uses to move from one state to another. This is necessary for the inclusion and convergence of thoughts, intentions, and perspectives in group cognition. While there is a large body of research on human intent recognition [50, 64] and cognitive state recognition [10], researchers have only recently acknowledged the importance of the *reciprocal* position, that humans need to understand the computer's reasoning. We review the topic only briefly here and refer the interested reader to chapters of this book that deal with it more directly.

Transparent or explainable machine learning is a topic of increasing interest. Stemming mainly from the need to support people who apply machine learning in, for example, health care [14] or finance [96], the need for understanding the internal states of machines is relevant also with regard to collaborative machines. Z.C. Lipton [61] points out, in opposition to popular claims, that simple models – such as linear

janin.koch@aalto.fi

models – are not strictly more interpretable than deep neural networks, because it depends on the notion of interpretability employed. The complexity of neural networks through different acting layers and raw input data increases the realism of presented results relative to human expectations; this supports interpretability of the machine's actions. In contrast, linear models rely on hand-engineered features, which can increase the algorithmic predictability but can render unexpected results, which are less expressible themselves.

T. Lei et al.'s approach of rationalising neural networks provides insight into the explainability of internal states on the basis of text analysis [58]. By training a separate neural network on subsections of the text, they highlighted those parts likely to have caused the decision of the main network. Another example of explaining deep neural networks is presented by L.A. Hendricks et al. [40]. They used a convolutional neural network to analyse image features and trained a separate recurrent neural network to generate words associated with the decision-relevant features. While this method provided good results, the explanatory power is tied to the structure of the network. In a third example, M.T. Ribeiro contributed his LIME framework, a technique to explain any classifier prediction, by learning a proxy interpretable model for certain locally limited predictions [68]. While the above-mentioned work focuses on the explainability of machine learning and AI output, a promising framework presented by T. Kulesza et al. describes some tenets for self-explainable machines [52]. In their work, a system was able to explain how each prediction was made and allowed the user to explain any necessary corrections back to the system, which then learned and updated in line with that input.

Most of today's approaches rely on separate training or manually added information, which limits the scope of these systems to carefully selected and limited tasks. In contrast, with more open-ended tasks, the potential context to be considered might not be manually pre-determined. We note that for group cognition it may not be necessary to explain to the user the reasoning that produced the outcome as opposed to a selected set of belief states. Their relevance is determined, in contrast, by the collaboration situation and the mental state of the communication partner. That poses a challenge for future work.

15.6.5 Predict Collaborative Actions of Others

Proceeding from their own knowledge and the reasoning of other group members, participants can predict others' behaviour. Again, this should be interpreted as a reciprocal process including *all* members of the group. While previous research has focused primarily on the prediction of human behaviour [67, 73], some recent work has looked at prediction of machine actions by a human [15].

Chandrasekaran et al. evaluated the modalities necessary to enable humans to create a 'Theory of AI Mind' [15]. In their study, participants were asked to infer the AI's answer with regard to a given image for questions such as 'How many people are in this image?', with or without additional information presented alongside the AI's

janin.koch@aalto.fi

response for the previous item. The users' answers could be seen as the behaviour expected of the AI. The modalities tested were a confidence barchart of the five best predictions; and an implicit and explicit attention map provided as a heatmap for the image. After the prediction, users were provided with instant feedback in the form of the system's answers. Users who had been presented with additional modalities too were shown to have results with equal or lower accuracy in comparison to users who received only the instant feedback. However, an increase in prediction accuracy after only a few trials indicates that users learned to predict the machine's behaviour better through familiarisation than via additional information about internal processes. The additional information seemed to encourage users to overadapt to system failures, which resulted in worse overall prediction. Further studies are needed to evaluate other potential sources of improved behaviour prediction. However, these first results might indicate that, to understand and predict AI, humans may need more information than that referring to reasoning alone.

The concept of group cognition comes from the discipline of collaborative learning, which has emphasised the necessity of each participant continuously learning and updating said participant's knowledge, concepts, and ideas. Having their origins in psychology, the notions behind collaborative learning assume human-level understanding, communication, and learning capabilities. In the context of collaborative *machines*, these traits do not exist yet and will have to be explicitly implemented. We will next consider some opportunities for such implementations.

15.6.6 Update Knowledge for Social Inference

During collaboration, the group members must integrate inferences of other participants with their existing knowledge. An extensive set of methods exists that may achieve this. Among these are inverse reinforcement learning [1], Bayesian belief networks [17], and variants of deep neural networks [22]. Results have been presented for social inference in special tasks, as language learning [17] and learning through presentation [1, 6]. However, these approaches assume for the most part that the human provides input for the machine to learn from, and they do not integrate the human more deeply into the loop.

Interactive machine learning adds the human to the loop but has mainly been applied for purposes of enriching data or boosting unsupervised or supervised learning [70]. P. Sinard et al. define interactive machine learning as machine learning wherein the user can provide information to the machine during the interaction process [80]. Meanwhile, A. Holzinger [43] considers interactive machine learning as a type of collaboration between algorithm and human [70]. He points out that not all input presented to a machine can be trained for, and that the machine has to be able to adapt to such situations. He presents an approach using an ant-colony algorithm to solve a travelling-salesman problem [44]. The algorithm presents the optimal path found thus far and allows the user to alter this path, in line with the contextual knowledge he possesses. Holzinger's results illustrate that this approach

janin.koch@aalto.fi

speeds up the discovery of the optimal path in terms of iteration when compared to machine-only optimisation. Even though this approach allows the machine and the human to work on a mutual goal, the common objective is fixed at the outset of the task.

Another line of research relevant in this context is that into multi-agent systems [84]. Work on multi-agent systems often refers to critical tasks such as disaster-response control systems [75] or autonomous cars [27], wherein the aim is of 'a mixture of humans performing high level decision-making, intelligent agents coordinating the response and humans and robots performing key physical tasks' [75]. For a review of multi-agent systems, we direct the reader to Y. Shoham and K. Leyton-Brown [78]. In general, research on human-in-the-loop multi-agent systems has focused on the task, the activity, and the role each agent should have in order to contribute to reaching the defined goal [12]. For example, A. Campbell and A.S. Wu highlight the criticality of role allocation, where a role is 'the task assigned to a specific individual within a set of responsibilities given to a group of individuals', for designing, implementing, and analysing multi-agent systems [12]. They further present computational models for various role-allocation procedures in accordance with a recent review of multi-agent methods. Role allocation, according to them, grows 'more important as agents become more sophisticated, multi-agent solutions become more ubiquitous, and the problems that the agents are required to solve become more difficult' [12]. While most multi-agent research looks at machine agents, as found in sensor–networks [4], some concepts and principles for the coordination of collaboration and for how roles within a group can be allocated in the most efficient way could be used for collaborative AI. However, in the strong sense of the word 'collaboration', most of the multi-agent methods do not foster interactive behaviour on common ground so much as favour individual task allocation. Nevertheless, experiences from these models can aid in understanding how roles influence this interaction.

15.6.7 Apply New Types of Initiative in Turn-Taking

While learning in groups is a shared task with a common goal, in open-ended interaction the goal depends on the current topics and can change as soon as new ideas start being explored. Hence, there is a need for understanding which knowledge most efficiently contributes to the current collaboration, and when. J. Allen et al.'s [5] well-known mixed-initiative interaction framework provides a method for inferring when to take initiative. Since Allen proposed it, this framework has been applied in various contexts of interactive systems, among them intelligent tutoring systems [37], interactive machine learning [16], and creative tools [26].

On the other hand, the decision on *what* to contribute presents a trade-off between context-aligned recommendations (following the current chain of thoughts) and exploratory recommendation (diversion from the current ideas). Contextually aligned reactions, analogously with value-aligned interactions [76], may take less

janin.koch@aalto.fi

effort to communicate and react to, for reason of existing context and already shared references. While these reactions are more likely to be understood by other group members, they do not necessarily explore new possible solution spaces. What could be called 'exploratory initiatives', on the other hand, bring with them the problem that future topics are partly unknown and that, accordingly, selection of the 'right' idea path to follow can be a thorny problem. This unknown solution space presents a challenge for selection, encouraging, and elaboration of new ideas. Perhaps the trade-off of initiatives that explore versus exploit new topics could be modelled in a manner paralleling that in optimisation. However, the first solutions for acting and learning in partly non-observable environments, known mainly as a partially observable Markov decision process (POMDP), are promising. Already, POMDPs are being used for decision-making and selection in human–robot interaction [49, 85, 95]. In T. Taha et al.'s work, for example, a POMDP guides the communication layer, which facilitates the flow and interpretation of information between the human and the robot [85]. Applying this information, the robot makes its action plan, while the current task, status, observed intention, and satisfaction are used to model the interaction within the POMDP. The paper's authors highlight that with a minimum amount of input the system was able to change the action plan or add corrective actions at any time.

While current research on interactive systems offer various approaches to coordinate, engage in, and facilitate interactions, none of them cover all the necessities for collaborative behaviour in the sense of group cognition. However, these approaches do present the prerequisites for future developments of such systems.

15.7 Conclusion

We have discussed cognitive abilities necessary for collaborative AI by building on the concept of group cognition. We reviewed some promising current approaches, which reflect that some aspects of these abilities are already identifiable and partially addressed. However, more research needs to be done. The main topics we have identified for future research are related to the expressiveness of machines, the ability to understand human interaction, and inherent traits of the behaviour of machines. We have highlighted in this context the necessity of extending and enhancing potential communication media of machines for purposes of more human-like communication, including social signal recognition within collaborative processes. Scholars researching collaborative machines could draw from previous experiences of human-robot interaction and adapt the findings to the particular context at hand. Another limitation of current approaches is related to the explainability of machine reasoning. In order to construct a 'Theory of AI Mind', as framed by Chandrasekaran et al. [15], a human has to be able to understand the reasoning behind an action, so as to recognise the machine's intent and most probable behaviour. We have presented several approaches to resolving this issue; however, the question of what best explains the reasoning of a machine remains. Finally, we must reiterate the necessity of extending current

janin.koch@aalto.fi

approaches in machine learning and interactive machine learning to act under the uncertainty conditions typical of human collaboration. This would enable machines to make suggestions and act in open-ended collaboration such as discussions and brainstorming, for which the idea space is not defined beforehand.

Acknowledgements The project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 637991).

References

1. Abbeel, P., Ng, A.Y.: Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the twenty-first international conference on Machine learning, p. 1. ACM (2004)
2. Abrams, D., Rutland, A., Palmer, S.B., Pelletier, J., Ferrell, J., Lee, S.: The role of cognitive abilities in children's inferences about social atypicality and peer exclusion and inclusion in intergroup contexts. Br. J. Dev. Psychol. **32**(3), 233–247 (2014)
3. Akkerman, S., Van den Bossche, P., Admiraal, W., Gijselaers, W., Segers, M., Simons, R.J., Kirschner, P.: Reconsidering group cognition: from conceptual confusion to a boundary area between cognitive and socio-cultural perspectives? Educ. Res. Rev. **2**(1), 39–63 (2007)
4. Alexakos, C., Kalogeras, A.P.: Internet of things integration to a multi agent system based manufacturing environment. In: 2015 IEEE 20th Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8. IEEE (2015)
5. Allen, J., Guinn, C.I., Horvitz, E.: Mixed-initiative interaction. IEEE Intell. Syst. Appl. **14**(5), 14–23 (1999)
6. Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A survey of robot learning from demonstration. Robot. Auton. Syst. **57**(5), 469–483 (2009)
7. Baker, M.J.: Collaboration in collaborative learning. Interact. Stud. **16**(3), 451–473 (2015)
8. Baker, M., Hansen, T., Joiner, R., Traum, D.: The role of grounding in collaborative learning tasks. Collab. Learn. Cogn. Comput. Approach. **31**, 63 (1999)
9. Bradáč, V., Kostolányová, K.: Intelligent tutoring systems. In: E-Learning, E-Education, and Online Training: Third International Conference, eLEOT 2016, Dublin, Ireland, August 31–September 2, 2016, Revised Selected Papers, pp. 71–78. Springer (2017)
10. Cai, Z., Wu, Q., Huang, D., Ding, L., Yu, B., Law, R., Huang, J., Fu, S.: Cognitive state recognition using wavelet singular entropy and arma entropy with afpa optimized gp classification. Neurocomputing **197**, 29–44 (2016)
11. Cambria, E., White, B.: Jumping nlp curves: a review of natural language processing research. IEEE Comput. Intell. Mag. **9**(2), 48–57 (2014)
12. Campbell, A., Wu, A.S.: Multi-agent role allocation: issues, approaches, and multiple perspectives. Auton. Agent. Multi-Agent Syst. **22**(2), 317–355 (2011)
13. Cannon-Bowers, J.A., Salas, E.: Reflections on shared cognition. J. Organ. Behav. **22**(2), 195–202 (2001)
14. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N.: Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1721–1730. ACM (2015)
15. Chandrasekaran, A., Yadav, D., Chattopadhyay, P., Prabhu, V., Parikh, D.: It takes two to tango: towards theory of ai's mind (2017). arXiv:1704.00717
16. Chau, D.H., Kittur, A., Hong, J.I., Faloutsos, C.: Apolo: making sense of large network data by combining rich user interaction and machine learning. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 167–176. ACM (2011)

janin.koch@aalto.fi

17. Cheng, J., Greiner, R.: Learning bayesian belief network classifiers: algorithms and system. In: Advances in artificial intelligence, pp. 141–151 (2001)
18. Chrislip, D.D., Larson, C.E.: Collaborative leadership: how citizens and civic leaders can make a difference, vol. 24. Jossey-Bass Inc Pub (1994)
19. Clark, H.H., Wilkes-Gibbs, D.: Referring as a collaborative process. Cognition **22**(1), 1–39 (1986)
20. Clark, H.H., Brennan, S.E., et al.: Grounding in communication. Perspect. Soc. Shar. Cogn. **13**(1991), 127–149 (1991)
21. Cohen, P.R., Perrault, C.R.: Elements of a plan-based theory of speech acts. Cogn. Sci. **3**(3), 177–212 (1979)
22. Dahl, G.E., Yu, D., Deng, L., Acero, A.: Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans. Audio Speech Lang. Process. **20**(1), 30–42 (2012)
23. Dartnall, T.: Artificial intelligence and creativity: an interdisciplinary approach, vol. 17. Springer Science & Business Media (2013)
24. de Haan, M.: Intersubjectivity in models of learning and teaching: reflections from a study of teaching and learning in a mexican mazahua community. In: The theory and practice of cultural-historical psychology, pp. 174–199 (2001)
25. De Jong, K.A., Spears, W.M., Gordon, D.F.: Using genetic algorithms for concept learning. Mach. Learn. **13**(2–3), 161–188 (1993)
26. Deterding, C.S., Hook, J.D., Fiebrink, R., Gow, J., Akten, M., Smith, G., Liapis, A., Compton, K.: Mixed-initiative creative interfaces (2017)
27. Dresner, K., Stone, P.: A multiagent approach to autonomous intersection management. J. Artif. Intell. Res. **31**, 591–656 (2008)
28. El Kaliouby, R., Robinson, P.: Mind reading machines: automated inference of cognitive mental states from video. In: 2004 IEEE International Conference on Systems, Man and Cybernetics, vol. 1, pp. 682–688. IEEE (2004)
29. El Kaliouby, R., Robinson, P.: Real-time inference of complex mental states from facial expressions and head gestures. In: Real-Time Vision for Human-Computer Interaction, pp. 181–200. Springer (2005)
30. Emojis as content within chatbots and nlps (2016). https://www.smalltalk.ai/blog/2016/12/9/how-to-use-emojis-as-content-within-chatbots-and-nlps
31. Engel, D., Woolley, A.W., Jing, L.X., Chabris, C.F., Malone, T.W.: Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. PloS one **9**(12), e115,212 (2014)
32. Flavell, J.H.: Theory-of-mind development: retrospect and prospect. Merrill-Palmer Q. **50**(3), 274–290 (2004)
33. Fotheringham, M.J., Owies, D., Leslie, E., Owen, N.: Interactive health communication in preventive medicine: internet-based strategies in teaching and research. Am. J. Prev. Med. **19**(2), 113–120 (2000)
34. Fussell, S.R., Kiesler, S., Setlock, L.D., Yew, V.: How people anthropomorphize robots. In: 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 145–152. IEEE (2008)
35. Galegher, J., Kraut, R.E., Egido, C.: Intellectual Teamwork: Social and Technological Foundations of Cooperative Work. Psychology Press (2014)
36. Goldstone, R.L., Theiner, G.: The multiple, interacting levels of cognitive systems (milcs) perspective on group cognition. Philos. Psychol. **30**(3), 334–368 (2017)
37. Graesser, A.C., VanLehn, K., Rosé, C.P., Jordan, P.W., Harter, D.: Intelligent tutoring systems with conversational dialogue. AI Mag. **22**(4), 39 (2001)
38. Gray, B.: Collaborating: Finding Common Ground for Multiparty Problems (1989)
39. Guzman, A.L.: The messages of mute machines: human-machine communication with industrial technologies. Communication+ 1 **5**(1), 1–30 (2016)
40. Hendricks, L.A., Akata, Z., Rohrbach, M., Donahue, J., Schiele, B., Darrell, T.: Generating visual explanations. In: European Conference on Computer Vision, pp. 3–19. Springer (2016)

janin.koch@aalto.fi

41. Hill, J., Ford, W.R., Farreras, I.G.: Real conversations with artificial intelligence: a comparison between human-human online conversations and human-chatbot conversations. Comput. Hum. Behav. **49**, 245–250 (2015)
42. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput.-Hum. Interact. (TOCHI) **7**(2), 174–196 (2000)
43. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. **3**(2), 119–131 (2016)
44. Holzinger, A., Plass, M., Holzinger, K., Crişan, G.C., Pintea, C.M., Palade, V.: Towards interactive machine learning (iml): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: International Conference on Availability, Reliability, and Security, pp. 81–95. Springer (2016)
45. Hong, H.Y., Chen, F.C., Chai, C.S., Chan, W.C.: Teacher-education students views about knowledge building theory and practice. Instr. Sci. **39**(4), 467–482 (2011)
46. Huber, G.P., Lewis, K.: Cross-understanding: implications for group cognition and performance. Acad. Manag. Rev. **35**(1), 6–26 (2010)
47. iOS Siri, A.: Apple (2013)
48. Jurafsky, D., Martin, J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2014)
49. Karami, A.B., Jeanpierre, L., Mouaddib, A.I.: Human-robot collaboration for a shared mission. In: Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, pp. 155–156. IEEE Press (2010)
50. Kelley, R., Wigand, L., Hamilton, B., Browne, K., Nicolescu, M., Nicolescu, M.: Deep networks for predicting human intent with respect to objects. In: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, pp. 171–172. ACM (2012)
51. Koch, J.: Design implications for designing with a collaborative ai. In: AAAI Spring Symposium Series, Designing the User Experience of Machine Learning Systems (2017)
52. Kulesza, T., Burnett, M., Wong, W.K., Stumpf, S.: Principles of explanatory debugging to personalize interactive machine learning. In: Proceedings of the 20th International Conference on Intelligent User Interfaces, pp. 126–137. ACM (2015)
53. Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B.: Human-level concept learning through probabilistic program induction. Science **350**(6266), 1332–1338 (2015)
54. Lala, D., Inoue, K., Milhorat, P., Kawahara, T.: Detection of social signals for recognizing engagement in human-robot interaction (2017). arXiv:1709.10257 [cs.HC]
55. Lang, F., Fink, A.: Collaborative machine scheduling: challenges of individually optimizing behavior. Concurr. Comput. Pract. Exp. **27**(11), 2869–2888 (2015)
56. Lave, J., Wenger, E.: Situated Learning: Legitimate Peripheral Participation. Cambridge university press, Cambridge (1991)
57. Lee, D., Lee, J., Kim, E.K., Lee, J.: Dialog act modeling for virtual personal assistant applications using a small volume of labeled data and domain knowledge. In: Sixteenth Annual Conference of the International Speech Communication Association (2015)
58. Lei, T., Barzilay, R., Jaakkola, T.: Rationalizing neural predictions (2016). arXiv:1606.04155
59. Levine, S.J., Williams, B.C.: Concurrent plan recognition and execution for human-robot teams. In: ICAPS (2014)
60. Licklider, J.C.: Man-computer symbiosis. IRE Trans. Hum. Factors Electron. **1**, 4–11 (1960)
61. Lipton, Z.C.: The mythos of model interpretability (2016). arXiv:1606.03490
62. Mavridis, N.: A review of verbal and non-verbal human-robot interactive communication. Robot. Auton. Syst. **63**, 22–35 (2015)
63. Mohammed, S., Ringseis, E.: Cognitive diversity and consensus in group decision making: the role of inputs, processes, and outcomes. Organ. Behav. Hum. Decis. Process. **85**(2), 310–335 (2001)
64. Nehaniv, C.L., Dautenhahn, K., Kubacki, J., Haegele, M., Parlitz, C., Alami, R.: A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. In: ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005, pp. 371–377. IEEE (2005)

janin.koch@aalto.fi

65. Novak, J.: Mine, yours... ours? Designing for principal-agent collaboration in interactive value creation. Wirtschaftsinformatik **1**, 305–314 (2009)
66. Oliver, N.M., Rosario, B., Pentland, A.P.: A bayesian computer vision system for modeling human interactions. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 831–843 (2000)
67. Pantic, M., Pentland, A., Nijholt, A., Huang, T.S.: Human computing and machine understanding of human behavior: a survey. In: Artifical Intelligence for Human Computing, pp. 47–71. Springer (2007)
68. Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you?: Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144. ACM (2016)
69. Rich, C., Ponsler, B., Holroyd, A., Sidner, C.L.: Recognizing engagement in human-robot interaction. In: 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 375–382. IEEE (2010)
70. Robert, S., Büttner, S., Röcker, C., Holzinger, A.: Reasoning under uncertainty: towards collaborative interactive machine learning. In: Machine Learning for Health Informatics, pp. 357–376. Springer (2016)
71. Robinson, T.N., Patrick, K., Eng, T.R., Gustafson, D., et al.: An evidence-based approach to interactive health communication: a challenge to medicine in the information age. JAMA **280**(14), 1264–1269 (1998)
72. Roschelle, J., Teasley, S.D., et al.: The construction of shared knowledge in collaborative problem solving. Comput.-Support. Collab. Learn. **128**, 69–197 (1995)
73. Ruttkay, Z., Reidsma, D., Nijholt, A.: Human computing, virtual humans and artificial imperfection. In: Proceedings of the 8th international conference on Multimodal interfaces, pp. 179–184. ACM (2006)
74. Sato, E., Yamaguchi, T., Harashima, F.: Natural interface using pointing behavior for human-robot gestural interaction. IEEE Trans. Industr. Electron. **54**(2), 1105–1112 (2007)
75. Schurr, N., Marecki, J., Tambe, M., Scerri, P., Kasinadhuni, N., Lewis, J.P.: The future of disaster response: humans working with multiagent teams using defacto. In: AAAI Spring Symposium: AI Technologies for Homeland Security, pp. 9–16 (2005)
76. Shapiro, D., Shachter, R.: User-agent value alignment. In: Proceedings of The 18th National Conference on Artificial Intelligence AAAI (2002)
77. Sheridan, T.B.: Human-robot interaction: status and challenges. Hum. Factors **58**(4), 525–532 (2016)
78. Shoham, Y., Leyton-Brown, K.: Multiagent systems: Algorithmic, game-theoretic, and logical foundations. Cambridge University Press, Cambridge (2008)
79. Sidner, C.L., Lee, C., Morency, L.P., Forlines, C.: The effect of head-nod recognition in human-robot conversation. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, pp. 290–296. ACM (2006)
80. Simard, P., Chickering, D., Lakshmiratan, A., Charles, D., Bottou, L., Suarez, C.G.J., Grangier, D., Amershi, S., Verwey, J., Suh, J.: Ice: enabling non-experts to build models interactively for large-scale lopsided problems (2014). arXiv:1409.4814
81. Soller, A.: Supporting social interaction in an intelligent collaborative learning system. Int. J. Artif. Intell. Educ. (IJAIED) **12**, 40–62 (2001)
82. Stahl, G.: Shared meaning, common ground, group cognition. In: Group Cognition: Computer Support for Building Collaborative Knowledge, pp. 347–360 (2006)
83. Stahl, G.: From intersubjectivity to group cognition. Comput. Support. Coop. Work (CSCW) **25**(4–5), 355–384 (2016)
84. Stone, P., Veloso, M.: Multiagent systems: a survey from a machine learning perspective. Auton. Robots **8**(3), 345–383 (2000)
85. Taha, T., Miró, J.V., Dissanayake, G.: A pomdp framework for modelling human interaction with assistive robots. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 544–549. IEEE (2011)
86. Theiner, G., Allen, C., Goldstone, R.L.: Recognizing group cognition. Cogn. Syst. Res. **11**(4), 378–395 (2010)

janin.koch@aalto.fi

87. Turner, P.: Mediated Cognition. Springer International Publishing, Cham (2016)
88. Vondrick, C., Oktay, D., Pirsiavash, H., Torralba, A.: Predicting motivations of actions by leveraging text. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2997–3005 (2016)
89. Vondrick, C., Pirsiavash, H., Torralba, A.: Anticipating visual representations from unlabeled video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 98–106 (2016)
90. Wenger, E.: Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge. Morgan Kaufmann (2014)
91. Wood, D.J., Gray, B.: Toward a comprehensive theory of collaboration. J. Appl. Behav. Sci. **27**(2), 139–162 (1991)
92. Woolf, B.P.: Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing e-Learning. Morgan Kaufmann (2010)
93. Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., Miyamoto, T.: Responsive robot gaze to interaction partner. In: Robotics: Science and Systems (2006)
94. Yu, Z., Ramanarayanan, V., Lange, P., Suendermann-Oeft, D.: An open-source dialog system with real-time engagement tracking for job interview training applications. In: Proceedings of IWSDS (2017)
95. Zhang, S., Sridharan, M.: Active visual sensing and collaboration on mobile robots using hierarchical pomdps. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1, pp. 181–188. International Foundation for Autonomous Agents and Multiagent Systems (2012)
96. Zhou, J., Chen, F.: Making machine learning useable. Int. J. Intell. Syst. Technol. Appl. **14**(2), 91–109 (2015)

janin.koch@aalto.fi