RIFL (2020) Vol. 14, n. 2: 76-87
DOI: 10.4396/AISB201907

From tools to social agents

Anna Strasser
Independent researcher
annakatharinastrasser@gmail.com

Abstract Up to now, our understanding of sociality is neatly tied to living beings. However, recent developments in Artificial Intelligence make it conceivable that we may entertain social interactions with artificial systems in the near future. With reference to minimal approaches describing socio-cognitive abilities, this paper presents a strategy of how social interactions between humans and artificial agents can be captured. Taking joint actions as a paradigmatic example, minimal necessary conditions for artificial agents are elaborated. To this end, it is first argued that multiple realizations of socio-cognitive abilities can lead to asymmetric cases of joint actions. In a second step, minimal conditions of agency and coordination in order to qualify as social agents in joint actions are discussed.

Keywords: artificial systems, social interaction, joint actions, minimal approach, social cognition

Received 25/01/2020; accepted 8/03/2020.

0. Introduction

Soon we will share a large part of our social lives with various kinds of artificial systems. Even though most interactions with artificial systems can sufficiently be described as tool use, it is worthwhile to investigate whether artificial agents may possess socio-cognitive abilities which enable them to overtake the role of a social agent and thereby constitute a new range of social interactions. Assuming that social interactions with artificial agents are conceivable, using artificial agents as tools for communicating with other human agents or retrieving information is not the only form of interaction. For example, think about future interactions with artificial agents designed to be companions such as care-robots and conversational machines (chatbots). Imagine an older person spending most of her time with a care-robot – not only using this robot as an assistant but also to communicate and to satisfy her social needs. I claim that categorizing such interactions as mere tool use will neglect essential social aspects. Having such interactions in mind, motivates to explore the potential role of artificial agents in the realm of social cognition and elaborate circumstances in which artificial agents could be considered as social agents and not as mere tools. However, our current conceptual framework concerning social agents seems to be restricted to living agents. Therefore, it cannot account for artificial agents as social agents. To overcome this restriction, we have two options: we can either propose an extension of our conceptions of tools, claiming that there are more complex tools which have some social features.

Alternatively, we can contemplate an extension of our conception of social agents. Assuming that there are interactions conceivable for which it is at least questionable whether we should classify them as tool use, I will pursue the latter.

1. Restrictive understanding of sociality

Up to now, our understanding of sociality is neatly tied to living beings. Social cognition is treated as a distinguishing feature of living beings. Research about social cognition includes topics such as social knowledge, social structure, group behavior, social influence, memory for social information, and attribution of motives (Frith, Blakemore 2006). All this is exclusively explored in living beings, such as humans and several species of the animal kingdom.

However, the picture changes a little bit if we focus on our practice of ascribing socio-cognitive abilities. We can distinguish two modes of ascriptions: an 'as if' mode and a justified mode of ascription. Justified ascriptions are restricted to living beings, whereas 'as if' ascriptions are frequently made to non-living beings. The 'as if' mode serves an explanatory role, its functional role consists in making sense of the world, but it remains neutral about the question of what socio-cognitive abilities objects really have. For instance, a famous experiment by Heider and Simmel (1944) illustrates how participants attribute social properties while describing simply moving geometrical forms. Although it is useful to characterize perceptual input through a social narrative and not through a technical description of geometric forms, it does not imply any ontological statement of whether described objects actually have social features. Along the same lines, Daniel Dennett (1987) describes how we apply the intentional stance to non-living beings.

Turning to standard philosophical notions in philosophy of mind characterizing socio-cognitive abilities as if they were unique to sophisticated adult human beings, we are even confronted with more restrictive notions. For example, the notion specifying individual agency (Davidson 1980) requires very demanding conditions concerning consciousness, goal generation, free choice, propositional attitudes, mastery of language, and intentionality. Likewise, the notion of joint action as introduced by Michael Bratman (2014) presupposes cognitively demanding conditions, such as having shared intentions requiring the ability to entertain a specific belief state, namely a relation of interdependence and mutual responsiveness, which in turn presupposes common knowledge. According to Bratman, only participants who are able to coordinate and build up explicit relations of commitment qualify as proper social agents in joint actions.

However, such notions account only for full-fledged, ideal cases. They cannot capture other forms of realizations with less demanding or simply different requirements. Research in developmental psychology, as well as in animal cognition, indicates that there are multiple realizations of socio-cognitive abilities. Moreover, even with respect to adult humans, one can observe that such ideal cases occur less often than expected. For instance, under time pressure otherwise sophisticated adults fall back on less demanding realizations. Likewise, developing expertise in a skill is often based on the automatization of formerly sophisticated processes.

Even though research indicates that there are multiple realizations of socio-cognitive abilities in various types of agents such as infants and non-human animals (Premack, Woodruff 1978; Heyes 2014, 2015) non-living beings still are in principle excluded from having social capacities. Therefore, the aim of considering non-living artificial agents as social agents presents a revolutionary challenge. In order to investigate how to account for sociality with respect to artificial agents, one has to explore how to overcome the restrictive nature of our current understanding of sociality.

Once we have a conceptual framework which is able to capture socio-cognitive abilities of artificial systems, new ethical questions concerning the consequences of potential social interactions will arise. Last but not least, analyzing potential consequences, one can evaluate whether developing artificial social agents is a desirable goal at all.

Inspired by the strategy of so-called minimal approaches (Butterfill, Apperly 2013; Michael *et al.* 2016) which offer notions for socio-cognitive abilities of infants and non-human animals, this paper discusses potential minimal necessary conditions with respect to artificial agents. The aim is to elaborate under which circumstances interactions with artificial agents qualify as social interaction even though we know that artificial agents are not living beings. When artificial systems prove to be capable of socio-cognitive abilities, this will constitute a new category of social interaction that still is reasonably similar to those we observe among humans or other living beings.

Recent studies in social neuroscience already demonstrate that interactions with artificial agents (avatars) are at least somehow comparable to interactions among humans. On the one side, social scientists study avatars as a way of understanding people (Scarborough, Bailenson 2014). Such studies investigate interactions between humans while they are embodied in avatars. Thereby special features of interacting with virtual representations are explored. On the other side, artificial agents are used in experimental designs in which participants are tricked in the sense that they believe that the interaction partner is a human counterpart while they are actually interacting with an artificial agent. If interactions with artificial systems would not have any similarities with human-human interactions, we could not use them to explore human behavior. However, it is important to note that this paper is not about working out to what extent people might be tricked by an artificial agent and attribute social characteristics in an 'as if' manner. Just taking an intentional stance (Dennett 1987) does not yet justify an attribution of socio-cognitive abilities as such. The aim here is to investigate whether artificial systems can actually have socio-cognitive characteristics. To explore this theoretical possibility, we have to be cautious not to mix 'as if' ascriptions with justified ascriptions. The question as to whether we are justified in ascribing socio-cognitive abilities to artificial systems is based on the assumption that the increasing experience of interacting with artificial agents is likely to alter our understanding of *social agents* radically.

2. Consequences of artificial social agents

The possibility that artificial agents might qualify as proper social agents in interactions with humans will raise new ethical and juristic questions. As soon as artificial agents can be understood as social agents, we need to pose questions concerning duties and rights those agents deserve as interaction partners.

There is a general agreement that artificial agents should not harm other living beings. This applies already to artificial agents which are considered as tools. However, as soon as we treat artificial agents as social agents (and not 'as if' they were social agents), ascribing duties will not be sufficient. At this point, one should consider whether this new type of a social agent also deserves rights. Since our self-understanding of fairness and justice is based on how we treat other social agents, it will be essential to develop social norms of how to treat artificial social agents. Already in a transitional phase, when artificial systems are not yet proper social agents, our interactions with them can influence our behavior towards other living social agents. Imagine a person using an artificial agent which is strikingly similar to a human in order to satisfy some felt needs, needs fulfillment which most people would find shameful, if not downright criminal, if it were to be acted out with another human agent. How can we preclude that this behavior with an artificial agent might not make it more probable that the person would end up crossing the line between fantasy and reality in the public world?

Furthermore, regarding the outcomes of joint actions in which artificial social agents are involved, we have to face new questions of responsibilities. For example, regarding responsibilities of autonomous driving systems, we might ask whether a person who is using an autonomous vehicle while having no way of taking over control, still should be held responsible for possible accidents (Hevelke, Nida-Rümelin 2015). The more autonomous artificial systems become, the more pressing becomes the question of whether simply the producers and the users are alone accountable for the outcome of the actions of those systems.

Where previous revolutions have dramatically changed our environments, this one has the potential to change our understanding of sociality and may lead to new social norms.

3. Socio-cognitive abilities of artificial agents

Accounting for socio-cognitive abilities of artificial agents, we have to assume that there are further multiple realizations that are not covered by our current conceptual framework. Recent research has already shown that there are certain multiple realizations of socio-cognitive abilities. For instance, we have data about how non-human animals and also very young infants are able to demonstrate social competences, which presumably are based on socio-cognitive abilities. Even though there are controversial debates about how exactly such competences are realized, it is obvious that the requirements of full-fledged, ideal cases are not fulfilled in such instances.

For example, it is common sense to assume multiple realizations regarding the ability to anticipate the behavior of others. Traditional conceptions of mindreading require the mastery of language as a necessary condition. But, the competence to anticipate the behavior of other agents has also been observed in populations where we cannot assume that a mastery of language is operating. Admittedly, interpretations of how this competence is realized in non-verbal populations are controversial. Some positions claim that those competences are best explained by the application of behavioral rules (Heyes 1998; Penn, Povinelli 2007; Whiten 2013) and thereby deny that ascribing mental states is part of such realizations. Others argue for genuine mindreading abilities (Fletcher, Carruthers 2013; Halina 2015). This debate is far from being decided; neither the behavioral nor the mentalistic interpretation can yet exclude the other. Consequently, the question as to whether infants or non-human animals possess the socio-cognitive ability of mindreading is still an open question. But undoubtedly there are multiple realizations of anticipating the behavior of others what I take as a socio-cognitive ability. Therefore, I argue that it is feasible to pose the question as to whether artificial systems can display socio-cognitive abilities.

Starting with the assumption that our understanding of socio-cognitive abilities is too restrictive, the exploration of minimal conditions for socio-cognitive phenomena concerning artificial systems will suggest an extension of our current conceptual framework for attributing socio-cognitive abilities.

4. Conditions for minimal joint actions

Joint actions constitute an interesting subset of social interactions, a subset in which people cooperate and do things together in order to reach a common goal. Taking joint actions as a paradigmatic example, this paper discusses minimal necessary conditions, which qualify artificial agents as proper participants in a joint action, and consequently as social agents. A minimal notion of joint action helps to distinguish tool use from joint actions and thereby enables a finer-grained description of human-computer interactions. Although the philosophical debate about joint actions is rather controversial, one can summarize some important requirements which are taken as necessary. Disagreements start when it comes to questions of sufficiency. An event can only qualify as a joint action if it is caused by the input of multiple agents. That means the effect of this event can be described as a common outcome of what several agents did, and whereby the *individual agencies* are intentional under some description (Davidson 1980). To distinguish mere plural activities from joint actions, we must require that both agents *aim* at bringing about the same effect. As a consequence, some sort of *coordination* is required. And it is further claimed that this coordination is achieved through special psychological mechanisms. However, the question as to whether these mechanisms can be based on shared goals (weak sense of joint action), or whether these mechanisms have to include shared intentions (strong sense of joint action) to ensure that not only the same goal is achieved but also that this goal is jointly aimed at, is still under debate.

Both strategies are problematic. The weak sense of joint action includes cases in which individual agents treat each other as social tools, whereas the stronger sense requires overly demanding conditions which are, for example, not fulfilled by young children. Inspired by Pacherie's notion of 'intention lite' (Pacherie 2013), I assume that there are middle cases which can exclude the social tool cases and at the same time refer to less demanding conditions.

To approach a solution, I first argue that there are *asymmetric cases of joint actions* in which the distribution of abilities is not equal among the participants. Uncontroversial cases of asymmetric joint actions are, for example, mother-child interactions. Despite the fact that infants do not fulfill the full-fledged sophisticated conditions of a strong sense of joint action, they are regarded as social agents in joint actions. That means they are able to act jointly with an adult participant while their fulfilled conditions differ from those of the adult. Consequently, it is sufficient to require less demanding conditions from one participant of a joint action. The performance of the participants in an asymmetric joint action can be based on multiple realizations. Consequently, artificial agents do not have to fulfill the very same conditions as required for human adults.

Since any notion of joint actions describes a plural activity, one has to presuppose the ability to act. Already at this point, we need an alternative notion of agency since standard philosophical notions (Davidson 1980) require rather human-centered abilities. In order to capture the notion of agency operative in artificial systems, we need a notion which does not rely on features we find only in biological systems. I have developed elsewhere a minimal notion of agency which does not rely on biological constraints (anonymized). This notion can capture artificial agents as potential actors. If artificial agents are not able to act in an appropriate sense, any further questions as to whether they might qualify as acting *jointly* would, of course, be a waste of time. For the sake of argument, I presuppose here that artificial systems can qualify as minimal actors. In line with the conception of asymmetric joint actions, a joint action performed by a mixed group of humans and artificial agents can then be seen as a combination of two types of agency.

The ability to coordinate will be at center stage in this investigation because coordination plays a crucial role for constituting the social dimension of joint actions. Regardless of whether one assumes shared goals or shared intentions, successful coordination in social interactions presupposes social competence. Agents must have some sort of an understanding of the other agents, which makes it possible to anticipate the other's behavior and to rely on the other's willingness to take over its part. Consequently, mindreading (or any other realization of anticipating behavior of others) and commitment are seen as important factors for ensuring the social competence needed for coordination, which is necessary in joint actions.

5. Anticipation – mindreading

It is common sense that a major function of social cognition consists in abilities to encode, store, retrieve, and process social information about conspecifics, as well as across species, in order to understand others. One crucial aspect, namely the ability to anticipate the behavior of other agents, plays an essential role in many social interactions. Being able to act jointly, we have to be able to anticipate what the other agent will do next. In the humanities and natural sciences, this aspect of social competence is discussed under the label of 'mindreading' or 'Theory of Mind' (Fodor 1992).

If artificial agents qualify as social agents in a joint action, we have to expect mindreading abilities from them. As we have already seen regarding the notion of agency, standard notions tend to be rather restrictive and demanding. The same is true for mindreading. Many conceptions of mindreading are tailored to adult humans and refer to a full-fledged form of mindreading requiring a mastery of language, as well as cognitively demanding abilities such as meta-representations.

Assuming multiple realizations and building upon minimal approaches, one can elaborate minimal necessary conditions of mindreading. In this paper, I am arguing that the less demanding conditions for *minimal mindreading* (Butterfill, Apperly 2013) provide an attractive alternative to capture the mindreading abilities of artificial agents.

In contrast to full-fledged mindreading, this approach specifies minimal presuppositions for mindreading. Instead of requiring a wide range of complex mental states, Butterfill and Apperly specify two mental states, namely encounterings and registrations. Roughly speaking, one may characterize encounterings as a kind of simple perception, whereas registrations could be described as a rudimentary form of believing. A minimal mindreader infers from observable cues to the mental state of encountering. With respect to the last observed encountering, the minimal mindreader ascribes a further mental state (registration) to the other agent. Finally, she applies a minimal theory of mind – which consists in the knowledge that goal-directed actions rely on registrations – to anticipate the behavior of the other. By representing encounterings and registrations, minimal mindreaders can in a limited but useful range of situations track others' perceptions and beliefs without representing perceptions and beliefs as such. Minimal mindreading is regarded as implicit, nonverbal, automatic, and is based on unconscious reasoning.

Research in artificial intelligence has already demonstrated that artificial agents can model mental states of human beings concerning the perspective of the human counterpart (Gray, Breazeal 2014). This shows that artificial agents, in principle, are able to infer from their perception of the physical world to what a human counterpart can see or cannot see in terms of an object. Furthermore, they can anticipate the behavior of the human counterpart as dependent on this perspective. For instance, they can take into account that the fact whether the human agent can see an object or not will guide her future actions. Insofar, we can claim that artificial agents succeed in some cases of minimal mindreading.

At this point, one can object that we are neglecting the genuine social aspect of mindreading. Admittedly, many examples in the mindreading debate tend to relate exclusively to mental states such as knowing and perceiving. Desires and emotions are not yet at the foreground of these debates. Focusing on the genuine social aspect, one can conclude that qualifying as a mindreader should include the ability to process social information. For instance, we do not only have to notice that another agent is noticing something relevant for the joint action, but we should also recognize whether the other agent is desiring something or is afraid of something. This presents a special challenge for artificial agents. Taking into account that human anticipatory systems fairly seamlessly include social and emotional aspects, we have to explore whether artificial systems are able to process social data as well. To anticipate future actions of other agents, it is not only relevant to consider their mental, but also their emotional states.

Turning to social data, actual research on social robotics is highly relevant, specifically in relation to the development of robots which are designed to enter the space of human social interactions. For example, research pertaining to conversational agents aims to develop artificial agents from mere tools into human-like partners (Mattar, Wachsmuth 2012; Becker, Wachsmuth 2006). Since the processing of social data plays an essential role in social interactions, I presuppose that artificial agents must be able to interpret the social cues presented by their interacting partners. In addition, social interactions are based on reciprocal exchanges. Therefore, artificial agents should also be able to send social cues in order to make their ‚minds' visible.

Much research is now focusing on social cues such as gestures (Kang *et al.* 2012) and emotional expression (Petta *et al.* 2011; Becker, Wachsmuth 2006). For example, ARIAs (Artificial Retrieval of Information Assistants) (Baur *et al.* 2015) are able to handle multimodal social interactions. They can maintain a conversation with a human agent and, indeed, they react adequately to verbal and nonverbal behavior. Even though results in social robotics may not yet apply to an unlimited range of situations, this shows that there are ways for artificial agents to process social data.

The above considerations indicate that artificial agents, in principle, are able to process social data and make use of it to anticipate the behavior of their interaction partners. Further developments in social robotics will probably also make it easier for the human counterpart to anticipate the behavior of the artificial agent.

However, according to traditional philosophical notions of mindreading, mere processing of emotional data is not taken as sufficient. In addition, having emotional and mental states is required. Assuming that mental or emotional states are exclusively found in living beings, our question as to whether artificial agents can be social interaction partners in a joint action turns into the question as to whether having mental and emotional states is a necessary requirement for realizing socio-cognitive abilities such as mindreading. One might argue that future AI systems might someday have mental and emotional states. But up until now, it does not look like as if this is to be expected in the near future. Therefore, the crucial question is whether we can ensure that we are not losing the sociality aspect even if we sacrifice mental and emotional states.

So far, the notion of minimal mindreading (Butterfill, Apperly 2013) is a promising starting point to characterize mindreading abilities of artificial agents. As we have seen, this notion questions the necessity of overly demanding cognitive resources, such as the ability to represent a full range of complex mental states and a mastery of language. And most importantly for artificial agents, the ability for minimal mindreading need not be based on conscious reasoning. Nevertheless, up to now, this notion has been only applied to living beings, only accounting for automatic mindreading in human adults, infants, and non-human animals. Even though this notion does not require conscious reasoning from a mindreading agent, future work will have to deliver further adjustments considering, for example, the processing of social data, before it can be applied to artificial systems.

In sum, one can argue that, in principle, artificial systems are able to process social and mental data and use it with a Theory of Mind to anticipate the behavior of human agents and thereby qualify as mindreaders. In a transition phase, it is likely that this works only in a very limited range of situations and it might be a special feature of asymmetric joint actions that they always only constitute a limited subset of joint actions.

6. Commitment

Another aspect of the required social competence enabling successful coordination in a joint action can be described as the ability to be committed to a joint action. To explore commitments regarding artificial agents, the recently developed notion of a minimal sense of commitment (Michael *et al.* 2016) presents a good starting point.

Commitments are relations between agents and an action which provide the security human social agents need to rely on each other. Additionally, commitments support the success of mindreading, since the behavior of agents who are sticking to their commitments is far easier to be predicted. In sum, one can claim that commitments function as the 'social glue' for much of what counts as social interactions.

Standard philosophical conceptions (Austin 1962; Searle 1969; Shpall 2014) characterize commitments as a relation between two or more agents and a specific action: An agent is committed to performing a specific action if she has assured her commitment and the other agent has acknowledged this. One component of a commitment is based on the motivation of one agent to contribute a specific action to a joint action; the other component is based on the corresponding expectation of the other agent that the counterpart will contribute to the joint action. Additionally, it can be claimed that this requires explicit acknowledgment and common knowledge. Standard conceptions of commitments rely on explicit utterances and are interpersonal since they describe a reciprocal relation between (at least) two agents. This can be contrasted with self-commitments which require just one agent.

Analyzing the possible classes of interpersonal commitments, it becomes obvious that standard conceptions neglect other potential cases. For example, not all interpersonal commitments require necessarily explicit assurances and acknowledgments. We experience implicit commitments in everyday life situations when agents feel and act committed even though no commitment was explicitly acknowledged (Gilbert 2006). Research in developmental psychology indicates cases of implicit commitments by showing that young children are capable of engaging in joint actions which rely on an interpersonal commitment without an explicit acknowledgment (Warneken 2006). Therefore, it seems uncontroversial to claim that commitments can also be realized in an implicit way.

Coming back to the notion of a minimal sense of commitment (Michael *et al.* 2016), we have a minimal approach to interpersonal commitments within which implicit commitments are also captured. It is of special interest with respect to the aim of this paper that this minimal approach additionally illuminates other neglected minimal forms of interpersonal commitments. Michael and colleges argue that components of a standard commitment, namely the expectation or the motivation, can be disassociated. Consequently, they claim that a single occurrence of just one component can be treated as a sufficient condition for a minimal sense of commitment. Presupposing that there is a goal of a potential joint action desired by one agent for which an external contribution of another agent is crucial, a minimal sense of commitment is already constituted if either one of the agents has a certain motivation, the other has a specific expectation, or both entertain the corresponding mental states.

In the standard cases, expectations are justified by the motivation of the other agent, whereas in minimal cases, the expectation of one participant can be sufficient. Applying this to asymmetric joint actions, a minimal sense of commitment realized by one participant (e.g., the human) can be sufficient. Assuming that artificial agents neither have emotional nor mental states displaying a minimal sense of commitment presents a real challenge for them. Future work will investigate whether artificial agents can display functionally equivalent states according to which it becomes reasonable to ascribe a minimal sense of commitment to them. However, concerning asymmetric joint actions, it is, for the most minimal case, sufficient if only human counterparts entertain a minimal sense of commitment.

7. Conclusion

Presupposing that artificial agents become increasingly prevalent in human social life, it is crucial to examine whether we are justified in ascribing socio-cognitive abilities to them and go on from there to consider artificial agents as social agents.

Starting with an examination of current and rather restrictive conceptions of sociality, this paper explored minimal necessary conditions enabling artificial agents to enter the realm of social cognition. One question was whether it could be a function of social cognition to encode, store, retrieve, and process social information not only concerning conspecifics or other species but also regarding artificial agents. Another question was whether artificial agents could have social cognition to encode, store, retrieve, and process social information concerning human beings.

Building upon multiple realizations of socio-cognitive abilities, I argued that there are asymmetric cases of joint actions in which the distribution of abilities is not equal among the participants. Therefore, artificial systems could take advantage of this asymmetry, so that, as has been argued, they do not have to fulfill the same – and idealized – conditions that are normally assumed to be fulfilled by living beings.

By easing the standard requirements for joint actions, which are based on demanding conceptions of agency and coordination, I suggested a minimal approach to joint actions to characterizing a joint action between artificial and human agents. In a first step, the demanding notion of agency was replaced by a minimal notion of agency according to which artificial systems can be seen, at least, as potential actors. In a second step, the presuppositions of successful coordination in joint actions were analyzed. The social competence to anticipate the behavior of other agents (mindreading) and to rely on their willingness to take over their part (commitment) were at the center of this investigation.

Developing minimal conditions for the requested social competence, I questioned whether having mental or emotional states is a necessary condition. Sacrificing mental or emotional states, it is crucial to ensure that we are not losing the sociality aspect, on which we are focusing when we discuss whether artificial agents qualify as social agents.

Concerning mindreading, a possible obstacle for artificial agents may be the ability to process and interpret social data such as gestures, facial expressions, and gaze following. However, developments in social robotics demonstrate that processing such social data is at least not impossible. It may not yet be sufficient to cover all sorts of social interactions, but it can cover a subset of social interactions. If having mental and emotional states is not a necessary requirement for successful processing of social data, a more completely developed notion of minimal mindreading (Butterfill, Apperly 2013) has the potential to capture the notion of social competence in artificial systems.

Focusing on the question as to under which circumstances a sense of commitment may arise in such interactions, considerations about the recently developed notion of a minimal sense of commitment (Michael *et al.* 2016) indicate how commitments can play a role in joint actions with mixed groups of artificial and human agents.

In sum, this sketch of a variety of minimal approaches describes joint actions of mixed groups of humans and artificial agents as a combination of two different sets of requirements. Whether interactions between two artificial agents may have social features will be a topic of future research.

In limited situations, we might even now claim that, for example, conversational machines are able to coordinate their speech acts to the speech acts of their dialogue partner, and thereby meet an important condition for joint action. Whatever future research will bring, with a conceptual framework that clarifies requirements for social agents, we can better characterize, understand, and regulate potential social interactions with artificial agents.

Bibliografia

Austin, J. (1962), *How to do Things with Words*, Harvard University Press, Cambridge (Mass.).

Baur, T., Mehlmann, G., Damian, I., Gebhard, P., Lingenfelser, F., Wagner, J., Lugrin, B., André, E. (2015), «Context-aware automated analysis and annotation of social human-agent interactions», in *ACM Trans. Interact. Intell. Syst.*, 5 (2), pp. 1-33.

Becker, C., Wachsmuth, I. (2006), «Modeling primary and secondary emotions for a believable communication agent», in *Proceedings of the 1st Workshop on Emotion and Computing in conjunction with the 29th Annual German Conference on Artificial Intelligence (KI2006)*, Bremen, pp. 31-34.

Bratman, M. (2014), *Shared agency: A planning theory of acting together*, Oxford University Press, Oxford.

Butterfill, S., Apperly, I. (2013), «How to construct a minimal theory of mind», in *Mind and Language*, 28 (5), pp. 606-637.

Davidson, D. (1980), *Essays on actions and events*, Oxford University Press, Oxford.

Dennett, D. (1987), *The Intentional Stance*, The MIT Press, Cambridge (Mass.).

Fletcher, L., Carruthers, P. (2013), *Behavior-Reading versus Mentalizing in Animals*, in Metcalfe, J., Terrace, H. (eds.) (2013), *Agency and Joint Attention,* Oxford University Press, Oxford, pp. 82-99.

Fodor, Jerry (1992), «Theory of the Child's Theory of Mind», in *Cognition,* 44 (3), pp. 283-296.

Frith, U., Blakemore, S.-J. (2006), *Social Cognition*, in Morris, R., Tarassenko, L., Kenward M. (eds.) (2006), *Cognitive Systems. Information Processing Meets Brain Science*, Elsevier Academic Press, Amsterdam, pp. 138-162.

Gilbert, M. (2006), «Rationality in collective action», in *Philos. Soc. Sci.*, 36, pp. 3-17.

Gray, J., Breazeal, C. (2014), «Manipulating Mental States Through Physical Action – A Self-as-Simulator Approach to Choosing Physical Actions Based on Mental State Outcomes», in *International Journal of Social Robotics*, 6 (3), pp. 315-327.

Halina, M. (2015), «There is no special problem of mindreading in nonhuman animals», in *Philosophy of Science*, 82(3), pp. 473-490.

Heider, F., Simmel, M. (1944), «An experimental study of apparent behavior», in *The American Journal of Psychology*, 57, pp. 243-259.

Hevelke, A., Nida-Rümelin, J. (2015), «Responsibility for crashes of autonomous vehicles: An ethical analysis», in *J. Sci Eng Ethics*, 21, p. 619.

Heyes, C. (1998), «Theory of Mind in Nonhuman Primates», in *Behavioral and Brain Sciences,* 21 (01), pp. 101-114.

Heyes, C. (2014), «False belief in infancy: a fresh look», in *Developmental Science*, 17 (5), pp. 647-659.

Heyes, C. (2015), «Animal mindreading: what's the problem?», in *Psychonomic Bulletin & Review*, 22 (2), pp. 313-327.

Kang, S., Gratch, J., Sidner, C., Artstein, R., Huang, L., Morency, L.P. (2012), «Towards building a virtual counselor: modeling nonverbal behavior during intimate self-disclosure», in *Eleventh International Conference on Autonomous Agents and Multiagent Systems*, Valencia, pp. 63-70.

Mattar, N., Wachsmuth, I. (2012), «Small talk is more than chit-chat: Exploiting structures of casual conversations for a virtual agent», in *KI 2012: Advances in artificial intelligence, Lecture notes in computer science,* vol. 7526, Springer, Berlin, pp. 119-130.

Michael, J., Sebanz, N., Knoblich, G. (2016), «The Sense of Commitment: A Minimal Approach», in *Frontiers in Psychology*, 6, 1968.

Pacherie, E. (2013), «Intentional joint agency: Shared intention lite», in *Synthese*, 190 (10), pp. 1817-1839.

Penn, D., Povinelli, D. (2007), «On the Lack of Evidence that Non-human Animals Possess Anything Remotely Resembling a 'Theory of Mind'», in *Philosophical Transactions of the Royal Society B*, 362 (1480), pp. 731-744.

Petta, P., Pelachaud, C., Cowie, R. (eds.) (2011), *Emotion-Oriented Systems: The Humaine Handbook*, Springer, Heidelberg.

Premack, D., Woodruff, G. (1978), «Does the chimpanzee have a theory of mind?», in *Behavioral Brain Sciences*, 1, pp. 515-526.

Scarborough, J., Bailenson, J. (2014), *Avatar Psychology*, in Grimshaw, Mark (ed.) (2014), *The Oxford Handbook of Virtuality,* Oxford University Press, Oxford, pp. 129-144.

Searle, J. (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge.

Shpall, S. (2014), «Moral and rational commitment», in *Philos. Phenomenological Res.*, 88 (1), pp. 146-172.

Warneken, F., Chen, F., Tomasello, M. (2006), «Cooperative activities in young children and chimpanzees», in *Child Dev.*, 77, pp. 640-663.

Whiten, A. (2013), «Humans are not alone in computing how others see the world», in *Animal Behaviour*, 86 (2), pp. 213-221.