Agency in Sociotechnical Systems: How to Enact Human–Robot Collaboration

Setareh Zafari, Sabine T. Koeszegi

Abstract

As artificial agents are introduced into diverse workplaces, basic configurations underlying the organization of work undergo a fundamental change. This implies that the work we do is subject to alteration along with who does the work that opens new social challenges. Questions regarding the extent of acceptance of these agents in work settings as well as the consequences of collaborating with artificial agents on human agents indicate the need to better understand the mechanisms that underpin a collaborative sociotechnical system. This book chapter discusses how the interplay between humans and artificial agents enables human–robot collaboration as a new way of working. Thus, we first focused on the agents and their interactive processes in the system to analyze how agency is ascribed to nonhuman entities. Thereafter, the results of two experiments are presented to reflect on the impact of attributing agency to an artificial agent on humans. This study provides recommendations for the design of artificial agents and organizational strategies in terms of which social practices and changes in the working context are required to provide possibilities for successful collaborations.

Keywords

Artificial agents, responsibility attribution, human–robot interaction, new ways of working

1 Agency in Sociotechnical Systems: How to Enact Human–robot Collaboration

Over the last decade, advances in the field of artificial intelligence (AI) have enabled passive objects to become active and agentic. Automation, which was principally focused on physical functions, has now begun to impact cognitive functions, such as complex motor coordination, perception processing, and decision-making [Hollnagel 1995]. An increasing use of assistant systems such as social robots and cobots, which are a specific form of robots that can work closely with humans [Faccio et al. 2019], is an example of artificial agents penetrating our everyday lives. Although a lot of the technology that allows these agents to initiate or respond to a variety of interactions with humans already exists, the social implications of these interactions enter new potentialities that have not been fully addressed.

Extensive research has shown that the subjective experience and willingness of humans to accept this integration is as relevant as the objective properties and functionalities of these technologies [De Graaf and Allouch 2013; Echterhoff et al. 2006]. Considering that humans are integral parts of these systems as sense-making actors and end-users, this book chapter analyzes the challenges associated with the integration of artificial agents into human social systems to facilitate collaboration.

https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_10

Developments in the field of AI suggest an increase in the autonomy of machines. One important implication of such autonomy is the ascribed (social) qualities, such as agency, which affects the perceptions and expectations of humans. The unpredictability of the actions undertaken by artificial agents may lead to situations where agency becomes an issue [Weber et al. 2013]. This unpredictability is mainly due to the inadequate understanding of computational mechanisms. Scholars have long debated the impact of attributing agency to machines on the diffusion of responsibility. For instance, Boos et al. [2013] argued that users can only be held accountable if they understand, predict, and influence work processes [Boos et al. 2013]. This approach emphasizes the fit of an actor's accountability demands and their control capabilities while considering the capacities of robots. Although attributing responsibility is the social function of being in control (i.e., a sense of agency) [Frith 2014], current artificial agents cannot be held responsible. Furthermore, understanding the notion of agency is relevant for improving user acceptance in human–machine interactions [Kim 2016; Lee et al. 2015], building trust in these relationships [Engen et al. 2016], and analyzing the ethical implications of smart technologies [Lin et al. 2012]. These arguments warrant a fresh look at artificial agents and their impact on human agents.

To analyze the challenges associated with collaboration with artificial agents, we focused on work settings and investigated how the interplay between humans and artificial agents enables collaboration. As the social is affected by material dimension but also affecting the material dimension [Leonardi 2012; Orlikowski 2009; Zammuto et al. 2007], it is necessary to study how this integration changes the sociotechnical dynamics of organizations. However, establishing the societal consequences of emerging forms of interaction with technology is beyond the scope of this chapter.

This chapter has been organized as follows. First, it defines the notion of agency and provides the framework for further analysis in sociotechnical systems, where humans are supported by technologies. Next, the results of two user experiments are discussed to provide a detailed picture of how collaboration with artificial agents affects humans and their basic needs. Finally, the contributions are summarized, and the directions of future research are discussed.

2 Enacting Human–robot Collaboration

To obtain a balanced emphasis on the social and technical aspects of working conditions, we used the sociotechnical system (STS) approach. This theoretical framework suggests that within organizations, humans (social) and technology (material) continually constitute the features of others. The notion of STS was developed by Trist and Bamforth in the mid-twentieth century to describe systems comprising a complex interaction between humans, machines, and the environmental aspects of the work system. Previously, the material dimension of organizations was considered as an external discrete input to the study of the social dimensions of organizations. Thus, it undermined the role of the social context in shaping the designs and uses of new technologies over time. However, this framework follows a relational ontology perspective [Law 2004; Barad 2007] and stresses the reciprocal interrelationship and entanglement of humans and technologies that shape technical and social working conditions. Although the social subsystem comprises individuals as members of the organization, the relationship among them, and their social attributes, the technical subsystem comprises the devices, techniques, and skills used by individuals to perform organizational tasks [Leonardi, 2012]. Thus, with a focus on the constitutive effect of the material and social dimensions, the properties of technologies and humans should be considered to explain how new affordances for working are created [Orlikowski 2009; Zammuto et al. 2007].

An underlying premise of this approach is that capacities for action are enacted in practice [Orlikowski and Scott 2008]. As machines become more sophisticated, understanding the agency attributed to these entities and its impact on humans and collaboration becomes even more critical. Agency is regarded as the capacity to act [Gray et al. 2007]. Two abstracted properties of agency are intentionality and autonomy [Bandura 1999; Banks 2019]. Intentionality is characterized by the capacity of an agent to process the contents of the mental state and justify actions or decisions. According to the theory of action [Davidson 1963], an action is intentional when it is caused by certain mental states. Thus, if no patterns of interaction and coordination based on expectations are identified, it is a coincidence and unintended. Autonomy is a combination of two Greek terms, auto (self) and nomos (governance) and is expressed in two dimensions: self-directedness (i.e., free will) and self-sufficiency (i.e., free act) [Bradshaw et al. 2013]. The former describes the agent's capability to take care of itself and create its agenda, while the latter describes the extent to which an agent is independent of external control. Thus, if no contingency or deviation from the set course is involved, an action is determined and preprogrammed.

Focusing on the agency of representative entities in a sociotechnical system can facilitate the development of more robust theories of the interrelationship between humans and artificial agents within a workplace. Moreover, it can potentially inform future strategic objectives for organizations that aim to integrate artificial agents. Therefore, this study analyzed how social and material entities and their agencies are continually coconstructed to enable a new way of working, namely human–robot collaboration. Human–robot collaboration refers to a collaborative partnership between humans and robots in completing tasks and focuses on coordinating joint activities between them [Ajoudani et al. 2018].

Collaboration can be differentiated from cooperation, where tasks for achieving a common goal are divided among participants, and each agent is responsible for only a part of the problem-solving. Collaboration is characterized by “the mutual engagement of participants in a coordinated effort to solve the problem together” [Roschelle and Teasley 1995, p. 70]. Therefore, collaborations require all agents to jointly engage in the entire task. That is, collaboration employs a complementarity approach and exceeds existing research that mostly substitutes humans with machines.

Studies have shown that human–AI collaboration can outperform a group of humans and sophisticated AI-based systems [Wang et al. 2016; Siegel 2016]. The resulting team success can be attributed to the unique advantages that emerge from combining human and AI capabilities in a compatible manner [Krüger et al. 2017]. Although the strengths of AI lie in analytical decision-making that involves the gathering and processing of large amounts of data, humans are well-versed in flexibility, creativity, and intuitive decision-making, particularly when heuristics are necessary for decision-making in uncertainty [Dragicevic et al. 2020; Jarrahi 2018]. Thus, artificial agents can extend human capabilities in task performance and decision-making.

To build an effective system, one needs to examine how integrating artificial agents reconfigures the main domains of an organization, including the i) division of labor and ii) integration of efforts. The former focuses on how to distribute tasks and decision rights among agents (human or artificial), and the latter elaborates on how to ensure the alignment of the efforts of different agents with the organizational goals. Therefore, studying the agents within this system is the first step to developing a better sense of the sociotechnical development process. However, using a sociotechnical approach to analyze collaboration with artificial agents does not mean categorizing social and technical actors and their actions but rather, showing the conditions of possibilities for these assumed categories or actors to behave in certain ways. Accordingly, it focuses on the flow of social formulations that enact those actions and performances [Hultin 2019]. Thus, we explored how agents (human and artificial) and their properties and identities are continuously performed to enable collaborative work between humans and robots as a new way of working.

Each subsection refers to an original work of the authors conducted as part of the dissertation. First, we focused on the agents and their interactive processes in the system to analyze how agency is ascribed to nonhuman entities (subsection 2.1). Thereafter, the results of two experiments are presented to reflect on the impact of attributing agency to an artificial agent on humans (subsections 2.2 and 2.3)

2.1. Robots as Artificial Agents

Several theoretical models such as the Actor–Network Theory [Latour 1996] and Double Dance of Agency [Rose and Jones 2005] suggest that ascribing agency is not limited to humans but also nonhuman entities, such as technologies. We differentiated between the agency of humans and that of machines and studied how these types of agencies are interrelated.

Recent developments in the field of AI suggest an increase in the agency of machines, as we assign them roles that were previously filled by humans. However, the unpredictability of the actions undertaken by artificial agents leads to situations where agency becomes an issue [Weber et al. 2013]. For instance, who would be responsible for the harm that is caused by a self-driving car? Considering that humans and machines do not possess the same capabilities [Engen et al. 2016; Rose and Jones 2005], we investigated the concept of agency and sought to comprehend the properties that humans seek when ascribing agency to nonhuman entities, such as robots.

Previous studies discovered different features related to our perception of machine agency, such as adaptability [Franklin and Graesser 1997], purposeful-looking movement [Scholl and Tremoulet 2000], complementary personalities [Lee et al. 2006], and humanlike appearance [Itoh and Inagaki 2004; Lee et al. 2015]. A seminal study in this area is the work of Rose and Turex [2000], which relates the perceived agency of machines to the human tendency toward anthropomorphism and describes machine agency as the extent to which machines are perceived by humans as having autonomy [Rose and Turex 2000].

We incorporated variable dimensions to develop a typology of artificial agents from a theoretical perspective. Typology is a conceptual classification that is mostly used in social, rather than natural sciences [Baily 1994]. It is one of the common styles of theorizing that systematically categorizes specific dimensions and features to create distinct types and profiles [Cornelissen 2017]. Classifying the artificial agents enables a deep and extended analysis of theories in previous studies about (social) agency to reflect on the possible consequences of human interactions with artificial agents on human–human interaction.

Depending on how machines control the input–output cycle and pursue the goal, we conceptualized artificial agents in four types, i) Non-AI marginally autonomous agents, ii) AI marginally autonomous agents, iii) AI semiautonomous agents, and iv) AI pseudoautonomous agents [Zafari and Koeszegi 2018]. A key distinction among these artificial agents is the extent to which they independently perform tasks. The autonomous consideration of AI marginally autonomous agents lies in their ability to move without human intervention, while AI semiautonomous agents adapt their goal settings because of their self-learning capacities. Responsibility implies autonomy; therefore, artificial agents are exempt from the usual responsibility practices and attribution. Thus, by finding such an artificial agent as the source of a failure or negative outcome, we need to understand how it determined the cause of failure to handle the issue and prevent a repetition. Thus, the responsibility for harm caused by artificial agents will always remain with human agents who initiate or manage the collaboration as the artificial agents are under the authority of the human agent in every step of the process.

The insights gained from this work [Zafari and Koeszegi 2018] may support the notion of collaborative agency [Kuziemsky and Cornett 2013]. Thus, agency does not belong to any actors and can be viewed as social affordance that emerges from the interaction between humans and artificial agents. This correlates with the relational ontology that argued that agency is constantly forming within the action [Law 2004; Barad 2007]. The agency attributed to an agent (human or artificial) may change in scale, over time and from one situation to another. Therefore, emphasis needs to be placed on the large-scale "system" at the heart of the analysis rather than discussing single agents to better elucidate organizational challenges.

2.2. Attitudes toward Artificial Agents

During collaboration, the activities of humans and robots occur in the same physical and social spaces [Dautenhahn and Sanders 2011]. This highlights the importance of the social aspects of interaction between these agents. Furthermore, ascribing agency to another entity highly depends on the physical and behavioral features of the entity and the characteristics of the perceiver [Takayama 2011; Waytz et al. 2010]. Studies on human–robot interaction (HRI) have mostly focused on the former [e.g., Itoh and Inagaki 2004; Lee et al. 2015; Lee et al. 2006], and there is still an extremely limited understanding of the cognitive processes that occur during HRI. Several technological features in robotics (such as increased sensitivity and safety) allow collaborative robots to support joint action in close contact with humans within a shared workspace [Bauer et al. 2008]. Inadequate effective management of social and cognitive features such as psychological safety [Edmondson 1999] and situational awareness [Cramton 2001] burden the collaboration between humans and robots. To provide insight into the cognition and intentional stance of humans while interacting with artificial agents, it is necessary to analyze the conditions that ensure the acceptance of the support of artificial agents without limiting human agency.

User studies about the Roomba robot showed that the owners exhibited different behaviors from the same robot vacuum cleaner; some gave it a name while it emptied its way, and some treated it like any other home application and did not talk to it [Takayama 2011; Forlizzi and Disalvo 2006]. This implied that the status of an entity's agency is not static, and the predefined and programmed functions of the entity and the perception of agency influence how we behave and interact with an entity. Moreover, recent studies [Appel et al. 2020; Złotowski et al. 2017] have shown that experience, as a dimension of mind perception, as well as agency, is related to an uncanny feeling toward humanlike robots and requires a better understanding of how ascribing agency elicits uncanniness or negative responses.

Although there is minimal theoretical knowledge regarding the agency of robots, it is necessary to not only describe but conduct an empirical study to explain under which conditions attributed agency positively/negatively impacts the attitudes toward robots. Thus, we conducted a vignette study and investigated the mechanism of the attitudes toward artificial agents. Vignettes refer to text, images, or videos that shortly describe a specific situation to evoke the attitudes or beliefs of participants concerning the present situation [Hughes and Huby 2002]. The flexibility of vignettes allows the exploration of factors and elements of interest by combining traditional survey and experimental design [Steiner et al. 2016]. Participants were asked to watch a video and respond to a postvideo questionnaire from the perspective of the vignette character as if they were that person in that situation.

We created two videos in which a human and robot collaborate to assemble a product (Figure 1). The main difference between the conditions is that under the "low agency" condition, the robot's behavior was relatively deterministic, while its behavior under the "high agency" condition was unpredictable. Thus, the actions of the cobot were not presuggested but were imperatively used to reflect the high level of autonomy.

The results showed that attributing high levels of agency to robots was associated with negative attitudes toward them only when individuals perceived low control during collaboration [Zafari and Koeszegi 2020]. Therefore, the lower the levels of decision control (inhibiting human autonomy), the lower the positive attitudes toward the robot with a high level of agency. Although preliminary, this finding highlights the role of the perception of control in promoting positive attitudes toward artificial agents. It implied that people do not perceive the high level of agency for artificial agents as negative except when they feel a lack of control during the work process. Furthermore, because perceived control is highly related to the diffusion of responsibility [Bandura 1991], it is necessary to consider the nature of perceived control in the collaborative context and establish approaches to enhance the perceptions of control for individuals working alongside artificial agents.

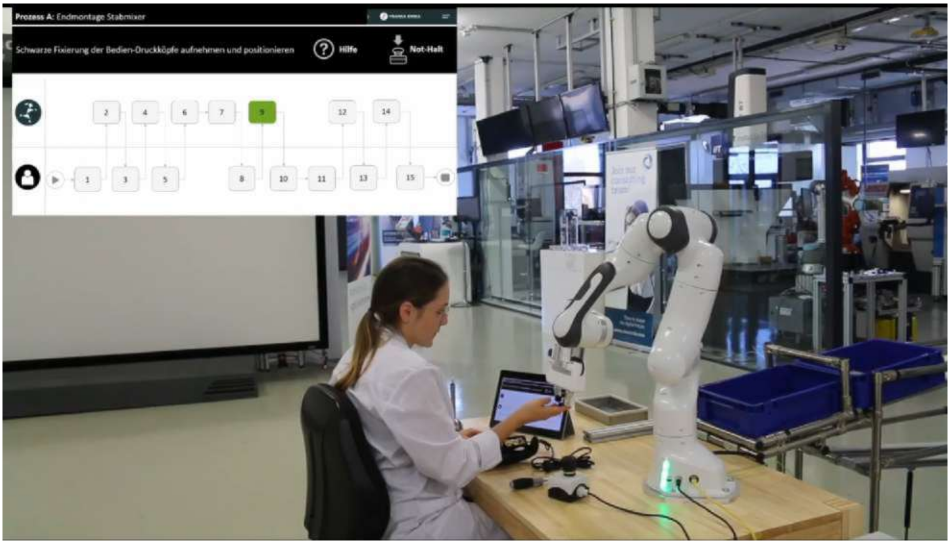

Figure 1 Screenshot of the video vignette that represents an artificial agent collaborating with a human agent

2.3. Interaction Style of Artificial Agents

A previous study on computers-are-social-actors established that social responses to computers fall under natural reactions to social situations; therefore, the principles drawn from sociology and social psychology are relevant for user interface design [Nass et al. 1994]. We interact with others according to our interpretation of the stimulus we receive from them [Blumer 1969]. The interpretation is a flexible social construct, which depends on the context and party involved [Pinch and Bijker 1984]. It helps us to clarify what to expect from the other party and is the basis for our future interactions.

Research predicts that service robots will soon be used within the social sphere of human agents as “natural” interaction partners [Floridi 2008]. With an increase in the entanglement of HRI, questions regarding the needs concerning the design of service robot applications have arisen. The appropriate design and implementation of robots serving with humans have been confirmed to be more challenging than old-fashioned industrial robots serving for humans. Robots serving for humans need to be capable of operating more or less autonomously and learning from errors, while robots serving with humans require the ability to communicate and interact with humans on a level involving understanding and responsiveness toward the human interaction partner [Kolbeinsson et al. 2019; Decker 2013]. This places a high demand on the quality of the interaction between humans and robots.

Considering that how a robot interacts with people can affect the efficiency of collaboration [Schulz et al. 2018], we focused on the interaction style of artificial agents and conducted a laboratory-based experiment with a Pepper robot developed by SoftBank Robotics using a built-in software. To design the interaction style of the robot, we referred to the “Big Two” dimensions of agency and communion [Bakan 1966]. Although the external validity of laboratory experiments is relatively lower than that of field experiments, they are a common method for HRI studies. A possible explanation for this is that most service robots are not easily accessible for daily usage since they are still in the research and development phase [Von der Puetten et al. 2018]. Laboratory experiments benefit from the high control over the extraneous variables that facilitate the replication of the conditions [Tanner 2018]. Therefore, they are useful for testing predictions and providing implication for designers of future robots.

We created two conditions of “person-oriented” and “task-oriented” interaction styles in which a service robot verbally assisted participants while they were building a house of cards. The robot under the person-oriented condition focused on socioemotional support and provided the participants with simple motivational phrases, while the robot’s focus under the task-oriented condition was on task performance and provided guidance concerning the goal and participant’s progress.

The experimental results showed that people interacting with a robot with a person-oriented interaction style reported higher self-efficacy in HRI, compared with that of a robot with a task-oriented interaction style. Moreover, we observed that several dimensions of the personality of a robot (specifically, extraversion, agreeableness, and emotional stability) can be simulated via robot verbal and speech interface design [Zafari et al. 2019]. These findings suggest the role of the interaction style of the robot in promoting perceived self-efficacy, which is crucial in developing trust in HRI. This implies that when a robot places emphasis on forming and maintaining a social relationship rather than pursuing goals and manifesting skills, an individual’s belief concerning their capabilities to perform in a particular situation heightens.

3 Discussion and Conclusion

Investigating the role of technology in organizations is a continuing concern within organizational research. Although new technologies are embraced for their capacity to create new ways of working, their disruptive impacts should not be undermined. This calls for a social science and human factor perspective to analyze the domains where these technologies potentially can and should be used and where they can but should not be used as their implementation may pose threats and challenges to organizations and society.

This book chapter discusses human–robot collaboration as a representative form of sociotechnical systems. It contributes to a better understanding of the impact of artificial agents on the behavior of human agents by discussing how the successful integration of the emerging technologies of AI and robotics in organizations depends not only on overcoming technical limitations but also considering social challenges. We demonstrated how the integration of artificial agents into social systems is reshaping the organization of work as the engagement with artificial agents creates the conditionality that makes certain practices enacted. Therefore, changes in work organization depend on assumed human agency, and the engagement with artificial agents creates a new arrangement of shared control in which agency is assigned and attributed to humans and artificial agents. This collaboration mindset helps to position human agents as the cocreators of the outcomes rather than the passive receiver of services provided by artificial agents. Thus, to better elucidate organizational challenges, we need to emphasize the system rather than the analysis of single agents.

The empirical findings reported in this chapter shed new light on social processes and their contribution to how people collaborate with artificial agents. As the ascribed agency to robots increases, the use of social processes in HRI also increases [Breazeal 2004]. The artificial nature of these agents presents several implications for their social interactions with humans; therefore, we suggested a set of contextual factors that influence the enactment of human–robot collaboration. We observed that the high perception of autonomy for an artificial agent leads to a lower acceptance and positive attitudes toward them when the level of perceived control for a human agent is low (inhibiting human autonomy). Furthermore, we observed that a robot's interaction style in providing feedback could be considered as a factor affecting self-efficacy in human collaborations. From a self-determination theory perspective, experiencing a sense of efficacy must be accompanied by a sense of autonomy [Deci and Ryan 2000] for intrinsic motivation to flourish, as the former resembles the need for competence, while the latter resembles the need for autonomy. Thus, the approach to the design of artificial agents requires the satisfaction of these human needs.

This emphasizes the importance of informal structure in enhancing the success of technological integration. The impact of delegating decisions and assigning roles to artificial agents in organizations is not limited to formal domains of organization (i.e., division of labor and integration processes) because the basic needs of individuals (i.e., needs for autonomy and competence), their work roles, and the social organizational structure are also affected. These findings suggest that for a successful integration of artificial agents into workspaces, a mindful consideration of the social components of interaction among humans and artificial agents is essential.

In addition to its exploratory nature, this chapter offers insight into which practices and changes in work organization are required to provide possibilities for successful integration. In this process, the key constructs are defined, the relationship between them is elucidated, and findings are discussed to demonstrate the viability of theoretical methods that offer minimal empirical support. This contribution can be classified as an intermediate theory [Edmondson and McManus 2007] that identifies new relationships among phenomena by reconceptualizing explanatory frameworks.

The scope of this study was limited in terms of work organization and analyzed how advances in the field of AI and robotics are affecting collaboration. A natural progression of this study is to analyze the possibilities and consequences of integrating these technologies into the tasks and processes that cannot yet be assigned to artificial agents, such as those requiring creativity. Further research can go beyond dyadic interactions between a human and artificial agent and explore how the team characteristics (such as the diversity or composition of a team) affect work dynamics and collaboration.

The collaboration process has a fundamental social component that robots working as the physical interaction partners of the human agent present a great risk on fundamental structures that are usually brought forth within human–human interaction, e.g., social norms. People expect artificial agents to apply the same norms that govern human–human interaction, and behavior that is not performed sufficiently similar to that of humans hinders the pragmatics of interaction [Sciutti et al. 2015]. Although humans will adapt to the capabilities of artificial agents [Hirschmanner et al. 2021] as well as the functionality of the sociotechnical system [Zafari et al. 2021], the impacts of constant interaction with artificial agents on the development and changes in social norms remain unclear. As Goffman [1983] emphasizes, the social self and individual actor are created through interactions [Goffman 1983]. The societal consequences of artificial agents penetrating the social lives of humans are intriguing and can be explored for further research.

Bibliography

Arash Ajoudani, Andrea Zanchettin, Serena Ivaldi, Alin Albu-Schaffer, Kazuhiro Kosuge, and Oussama Khatib. 2018. Progress and prospects of the human–robot collaboration. *Autonomous Robots* 42, 5 (2018), 957–975.

Markus Appel, Silvana Weber, Stefan Krause, and Martina Mara. 2016. On the eeriness of service robots with emotional capabilities. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction.* 411–412.

Kenneth Bailey. 1994. *Typologies and taxonomies: An introduction to classification techniques.* 102. Sage.

David Bakan. 1966. *The duality of human existence: Isolation and communion in western man.* Beacon Press.

Albert Bandura. 1991. Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes* 50, 2 (1991), 248–287.

Albert Bandura. 1999. Social cognitive theory: An agentic perspective. *Asian journal of social psychology* 2, 1 (1999), 21-41.

John Banks. 2019. A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior* 90 (2019), 363–371.

Karen Barad. 2007. *Meeting the Universe Halfway: Quantum physics and the entanglement of matter and meaning.* Duke University Press.

Andrea Bauer, Drik Wollherr, and Martin Buss. 2008. Human–robot collaboration: A survey. *International Journal of Humanoid Robotics* 5, 1 (2008), 47-66.

Herbert Blumer. 1969. *Symbolic interactionism: Perspective and method.* Univ. of California Press.

Daniel Boos, Hannes Guenter, Gudela Grote, and Katharina Kinder. 2013. Controllable accountabilities: The Internet of Things and its challenges for organisations. *Behaviour and Information Technology* 32, 5 (2013), 449–467.

Jeffery Bradshaw, Robert Hoffman, David Woods, and Matthew Johnson. 2013. The seven deadly myths of "autonomous systems." *IEEE Intelligent Systems* 28, 3 (2013), 54–61.

Robert Siegel. 2016. 20 Years later, humans still no match for computers on the chessboard. *NPR*. Available at https://www.npr.org/sections/alltechconsidered/2016/10/24/499162905/20-years-later-humans-still-no-match-for-computers-on-the-chessboard

Joep Cornelissen. 2017. Editor's comments: Developing propositions, a process model, or a typology? Addressing the challenges of writing theory without a boilerplate. *Academy of Management Review* 42, 1 (2017), 1-9.

Catherine Cramton. 2001. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science* 12, 3 (2001), 346–371.

Kerstin Dautenhahn, and Joe Sanders. 2011. Introduction. In *New frontiers in human-robot interaction,* Dautenhahn, K, and Sanders, J (Eds.). John Benjamins Publishing. 1–5.

Donald Davidson. 1963. Actions, Reasons, and Causes, *Journal of Philosophy* 60, 23 (1963), 685-700.

Edward Deci, and Richard Ryan. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11, 4 (2000), 227-268. DOI:10.1207/S15327965PLI1104_01

Maartje De Graaf, and Somaya Ben Allouch. 2013. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems* 61, 12 (2013), 1476–1486.

Michael Decker. 2013. Robotik. In *Handbuch Technikethik,* Grunwald, A, Simonidis-Puschmann, M (Eds.). Stuttgart JB Metzler, 354–358, https://doi.org/10.1007/978-3-476-05333-6_67

Gerald Echterhoff, Gerd Bohner, and Frank Siebler. 2006. "Social Robotics" und Mensch-Maschine-Interaktion. *Zeitschrift Für Sozialpsychologie* 37, 4 (2006), 219–231.

Amy Edmondson, and Stacy McManus. 2007. Methodological fit in management field research. *Academy of Management Review* 32, 4 (2007), 1246–1264.

Vegard Engen, J Brian Pickering, and Paul Walland. 2016. Machine agency in human-machine networks; impacts and trust implications. *Lecture Notes in Computer Science*, 9733. Springer, Cham. 96–106.

Maurizio Faccio, Matteo Bottin, and Giulio Rosati. 2019. Collaborative and traditional robotic assembly: A comparison model. *The International Journal of Advanced Manufacturing Technology* 102, 5 (2019), 1355–1372.

Luciano Floridi. 2008. Artificial intelligence's new frontier: Artificial companions and the fourth revolution. *Metaphilosophy* 39 (2008), 651–655.

Jodi Forlizzi, and Carl DiSalvo. 2006. Service robots in a domestic environment: A study of the Roomba vacuum in the home. In *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM. https://doi.org/10.1145/1121241.1121286

Stan Franklin, and Art Graesser. 1997. Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents. In: Müller J P, Wooldridge M J, Jennings N R (eds) *Intelligent Agents III Agent Theories, Architectures, and Languages. ATAL 1996. Lecture Notes in Computer Science,* vol 1193. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0013570

Chris Frith. 2014. Action, agency and responsibility. *Neuropsychologia* 55, 1 (2014), 137–142.

Heather Gray, Kurt Gray, and Daniel Wegner. 2007. Dimensions of mind perception. *Science* 315, 5812 (2007), 619. https://doi.org/10.1126/science.1134475

Matthias Hirschmanner, Stephanie Gross, Setareh Zafari, Brigitte Krenn, Friedrich Neubarth, and Markus Vincze. 2021. Investigating Transparency Methods in a Robot Word-Learning System and Their Effects on Human Teaching Behaviors. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).* 175–182. https://doi.org/10.1126/science.1134475

Erik Hollnagel. 1995. Cognitive functions and automation: principles of human-centred automation. *Advances in Human Factors/Ergonomics* 20 (1995), 971–976.

Rhidian Hughes, and Meg Huby. 2002. The application of vignettes in social and nursing research. *Journal of Advanced Nursing* 37,4 (2002), 382–386.

Lotta Hultin. 2019. Information and organization on becoming a sociomaterial researcher: Exploring epistemological practices grounded in a relational, performative ontology. *Information and Organization* 29, 2 (2019), 91–104.

Makoto Itoh, and Toshiyuki Inagaki. 2004. A microworld approach to identifying issues of human-automation systems design for supporting operator's situation awareness. *International Journal of Human-Computer Interaction* 17, 1 (2004), 3–24.

Mohammad Jarrahi. 2018. Artificial intelligence and the future of work: *Human-AI symbiosis in organizational decision making. Business Horizons* 61, 4 (2018), 577–586.

Katherine Kim. 2016. Interacting socially with the internet of things (IoT): Effects of source attribution and specialization in human-IoT interaction. *Journal of Computer-Mediated Communication* 21,2 (2016), 420–435.

Ari Kolbeinsson, Erik Lagerstedt, and Jessica Lindblom. 2019. Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing human-robot cooperation in manufacturing. *Production and Manufacturing Research* 7, 1 (2019), 448–471.

Craig Kuziemsky, and Janet Cornett. 2013. A model of collaborative agency and common ground. *In ITCH.* IOS Press. 388–392.

Bruno Latour. 1996. On actor-network theory: A few clarifications. *Soziale Welt*, 369-381.

James Law. 2004. *After method: Mess in social science research.* Routledge.

Jae-Gil Lee, Ki Joon Kim, Sangwon Lee, and Dong-Hee Shin. 2015. Can autonomous vehicles be safe and trustworthy? Effects of appearance and autonomy of unmanned driving systems. *International Journal of Human-Computer Interaction,* 31(10), 682–691. https://doi.org/10.1080/10447318.2015.1070547

Kwan Lee, Wei Peng, Seung Jin, and Chang Yan. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication* 56, 4 (2006), 754–772.

Paul Leonardi. 2012. Materiality, sociomateriality and socio-technical systems: What do these terms mean? How are they related? Do we need them? In Leonardi, P M, Nardi, B A, Kallinikos, J (Eds.), *Materiality and organizing: Social interaction in a technological world.* Oxford University Press. 25-48.

Patrick Lin, P., Keith Abney, and George Bekey. 2012. *Robot Ethics: The social and ethical implications of robotics.* MIT Press, MA2012.

Clifford Nass, Jonathan Steuer, and Ellen Tauber. 1994. Computers are social actors. *CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 72–78. https://doi.org/10.1145/191666.191703

Wanda Orlikowski, and Susan Scott. 2008. Sociomateriality: Challenging the separation of technology, work and organisation. *The Academy of Management Annals* 2, 1 (2008), 433-474.

Wanda Orlikowski. 2009. The sociomateriality of organisational life: considering technology in management research. *Cambridge Journal of Economics* 34, 1 (2009), 125–141.

Trevor Pinch, and Wiebe Bijker. 1984. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science* 14, 3 (1984), 399–441.

Jeremy Roschelle, and Stephanie Teasley. 1995. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning,* Springer. 69–97.

Jeremy Rose, and Matthew Jones. 2005. The double dance of agency: A socio-theoretic account of how machines and humans interact. *An International Journal on Communication, Information Technology and Work* 1, 1 (2005), 19–37.

Brian Scholl, and Patrice Tremoulet. 2000. Perceptual causality and animacy. *Trends in Cognitive Sciences* 4, 8 (2000), 299–309.

Ruth Schulz, Philipp Kratzer, and Marc Toussaint. 2018. Preferred interaction styles for human-robot collaboration vary over tasks with different action types. *Frontiers in Neurorobotics,* 12 (2018), 36.

Alessandra Sciutti, Caterina Ansuini, Cristina Becchio, and Giulio Sandini. 2015. Investigating the ability to read others intentions using humanoid robots. *Frontiers in Psychology,* 6 (2015), 1–6.

Peter Steiner, Christiane Atzmüller, and Dan Su. 2016. Designing valid and reliable vignette experiments for survey research: A case study on the fair gender income gap. *Journal of Methods and Measurement in the Social Sciences,* 7, 2 (2016), 52–94.

Leila Takayama. 2012. Perspectives on agency interacting with and through personal robots. In *Human-computer interaction: the agency perspective.* Springer, Berlin, 195–214

Kerry Tanner, 2018. Experimental research. In *Research Methods.* Chandos Publishing. 337–356

Astrid M von der Pütten, Nicole C Krämer, Cristina Becker-Asano, Kohei Ogawa, Suichi Nishio, and Hiroshi Ishiguro. 2018. At the café—Exploration and analysis of people's nonverbal behavior toward an android. In *Geminoid Studies,* Ishiguro H, Dalla Libera F (eds). Springer, 375–397. https://doi.org/10.1007/978-981-10-8702-8_24

Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew Beck. 2016. Deep learning for identifying metastatic breast cancer. *arXiv preprint* arXiv:1606.05718

Adam Waytz, Kurt Gray, Nicholas Epley, and Daniel Wegner. 2010. Causes and consequences of mind perception. *Trends in Cognitive Sciences*, 14, 8 (2010), 383–388.

Rolf Weber, Angela Pereira, Francien Dechesne, Job Timmermans, Rob van Kranenburg, and Hendrik Lehn. 2013. *Fact sheet - Ethics Subgroup IoT - Version 4. 0.* Delft University of Technology Chair Ethics Subgroup IoT Expert Group, 1–21.

Setareh Zafari, Sabine Koeszegi, Michael Filzmoser. 2021. Human Adaption in the Collaboration with Artificial Agents. In *Konnektivität Über die Bedeutung von Zusammenarbeit in der virtuellen Welt*, J Fritz, N Tomaschek (Eds.). Waxmann Verlag GmbH, Münster, Deutschland, (2021), 97–106.

Setareh Zafari and Sabine Koeszegi. 2020. Attitudes toward attributed agency: Role of perceived control. *International Journal of Social Robotics.* 13 (2020), 2071–2080. https://doi.org/10.1007/s12369-020-00672-7

Setareh Zafari, Isabel Schwaninger, Matthias Hirschmanner, Christina Schmidbauer, Astrid Weiss, and Sabine Koeszegi. 2019. "You Are Doing so Great!" – The Effect of a Robot's Interaction Style on Self-Efficacy in HRI. In *proceedings of IEEE International Symposium on Robot and Human Interactive Communication (ROMAN).* pp. 1–7, doi: 10.1109/RO-MAN46459.2019.8956437.

Setareh Zafari and Sabine Koeszegi. 2018. Machine agency in socio-technical systems: A typology of autonomous artificial agents. In *Proceeding of 2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO).* Doi: 10.1109/ARSO.2018.8625765.

Raymond Zammuto, Terri Griffith, Ann Majchrzak, Daniel Dougherty, and Samer Faraj. 2007. Information technology and the changing fabric of organization. *Organization Science* 18, 5 (2007), 749–762

Jacob Złotowski, Kumar Yogeeswaran, and Christoph Bartneck. 2017. Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies.* 100(2017), 48–54.