Understanding Emotions and Their Significance through Social Robots, and Vice Versa

Guest Editors' Introduction

Johanna Seibt and Raffaele Rodogno

Social robots are often generically characterized as 'artificial embodied agents that can act autonomously in the physical and symbolic space of human social interactions.' The particular philosophical interest of these artefacts more clearly comes into view, however, once their purposes and idiosyncratic effects are revealed: Social robots are a new class of agents—embodied artificial agents that fulfil tasks and afford human experiences that so far *only* the interaction with other people and animals could provide.

What are we to make of this fact—in conceptual-descriptive and in normative-practical regards? What does the idea of 'social' robots imply for our conceptions of sociality, and should we promote or discourage such interactions? These questions are thrown into stronger relief as soon as we turn to yet another characterization of this new class of artificial 'social' agents that highlights the underlying concrete design strategies:

> Ideally, people will treat [the social robot] Kismet as if it were a socially aware creature with thoughts, intents, desires, and feelings. Believability is the goal. Realism is not necessary. (Breazeal 2002, 52)

> Social robots are built to be expressive in ways that mimic the expressivity of living beings. By tracking human movement, lowering their eyes, tilting their heads, and other such movements, they promote emotional at-

Johanna Seibt, Research Unit for Robophilosophy, School of Culture and Society, Aarhus University, Jens-Christian Skousvej 7, 8000 Aarhus C, Denmark; filseibt@cas.au.dk.

Raffaele Rodogno, Research Unit for Robophilosophy, School of Culture and Society, Aarhus University, Jens-Christian Skousvej 7, 8000 Aarhus C, Denmark; filrr@cas.au.dk.

tachments and a desire to interact in their users. (Van Grunsven and Van
Wynsberghe 2019, 300)

If the 'social awareness' of such artificial agents includes, in the eyes of the ro-
boticist, "desires and feelings," and if the bodily expressiveness of such artificial
agents is successfully exploited to create (intended and unintended) emotional
reactions when people experience themselves as having 'social' interactions with
such robots, it appears that the 'question of emotions' is of pivotal significance in
exploring the normative and the conceptual-theoretical tasks that arise with social
robotics.

Human-Robot Interaction research (officially abbreviated as 'HRI') and re-
search in Social Robotics have been studying the emotional dimension of human
interactions with robots in empirical regards for over two decades (see, e.g., Fel-
lous and Arbib 2005; Novikova and Watts 2014; Kolling et al. 2016; Jerčić et
al. 2018), as well as exploring different strategies in "affective computing" and
in constructing "emotion architectures" (see, e.g., Vallverdú 2009; Saldien et al.
2010; Scheutz 2014; Lisetti and Hudlicka 2015). The philosophical debate about
social robots so far has focused on the possible ethical implications of using such
technology broadly in society, and in this connection the fact that robots (are built
to) evoke human emotions has been a frequent *topos*. However, surprisingly, the
topic of emotions *as such* has not been in the spotlight of the philosophical reflec-
tion on social robots.

This circumstance can be traced to two facts. First, emotions are much less
well understood in (western) philosophy than thoughts or even (self-) conscious-
ness—only since the 1980s has the 'philosophy of emotion' become a focused
research area, notwithstanding important single investigations throughout the his-
tory of philosophy. This fact is reflected in a persistent terminological uncertainty,
not only in philosophy but also other pertinent disciplines such as psychology,
cognitive science, anthropology, and robotics, where the terms 'emotion,' 'feel-
ing,' 'affect,' and 'experienced bodily sensation' are used in different and often
incongruent ways.

Second, and perhaps even more importantly, philosophical definitions of
social relations and social interactions have concentrated on the analysis of the
normative significance of the peculiar mutuality of these relations, the structure of
commitments incurred, and the formation conditions for such relations and inter-
actions at the conceptual-cognitive level (e.g., whether social interactions require a
'theory of mind'). The emotional dimension of social interactions is often not even

mentioned in the 'classical' sources on social action, interaction and joint action (see, e.g., Gilbert 1992; Miller 2001; Searle 2009; Tuomela 2012), since it is not taken to be a constitutive factor.

The phenomena investigated in HRI and social robotics research suggest, however, that emotions deserve a different role in the philosophy of sociality and social ontology. They open up a new perspective on social cognition and social action that links embodiment, emotion, and commitment in ways that are of central relevance for both theoretical and practical philosophy. The aim of this special issue is thus threefold.

On the one hand, at a general level, i.e., as a contribution to a new research landscape in formation, we wish to draw attention to the fact that the philosophical reflection on the phenomena of social robotics not only has a practical but also an important theoretical dimension. In fact, the new class of embodied artificial agents with capacities for symbolic expression and (the simulation of) social competence challenges the basic assumptions of our traditional (Cartesian) conception of subjectivity—the assumption that self-consciousness, rationality, individuality, freedom, agency, normative competence, responsibility, and moral dignity are a package deal. As we and our colleagues at the "Research Unit for Robophilosophy" at Aarhus University have argued since 2014, the phenomena of artificial 'social' agency raise new questions for all areas of philosophy. Beyond "robo-ethics" and "machine ethics," we need a more comprehensive response in the form of "robophilosophy"—"philosophy of, for, and by social robotics" (Seibt 2017a)— which is a new area of experimental and interdisciplinary applied philosophy.

On the other hand, the issue is to contribute to the more specific theoretical question about the role of emotions in the philosophy of mind and social ontology. We hope that the contributions collected here will serve to highlight that emotions are not merely a byproduct of social interactions that conceptual analyses of sociality can bypass, but rather play a pivotal role for our understanding of the processes of social cognition, both pre-conscious and within the scope of awareness, and for the modeling of the connection between social cognition and social interaction. The phenomena of human-robot interaction convey in striking ways that the linkages between motion perception, emotional response, sociality attributions and commitment deserve much closer attention and analysis than the current practice of shoving them under the label of "anthropomorphization" without reflecting details and possible differentiations.

Finally, as the contributions to the issue illustrate, within the context of social robotics the topic of emotions intertwines conceptual, empirical, and ethical

perspectives. This means in particular that the common methodologies of applied ethics cannot be straightforwardly applied. Ethical evaluations of social robotics applications not only depend on familiarity with the available empirical research, as elsewhere in applied ethics, but are complicated by the novelty of human-robot interactions and the lack of a clear and precise conceptualization of these phenomena. The contributions to this special issue will be helpful, we hope, in interfacing the ethical debate not only with empirical research on social cognition but also with the investigations in theoretical philosophy, especially philosophy of mind conducted in the paradigm of embodied cognition or interactivism, as well as (social) phenomenology, to come to grips with the conceptual questions of artificial embodied sociality.

As the contributions to this special issue show, however, putting the spotlight on emotions serves well to focus the ramified questions of social robotics onto the following two core issues:

1. How does the expressivity of the human body (in various senses of the term) relate to the communication (expressing and eliciting) of emotions, and what does it betoken for our understanding of human social cognition and social action that this expressivity can be simulated in robots, to a certain degree?

2. Which normative—instrumental, ethical, or more widely socio-cultural— reasons do we have to endorse or suppress such communicative (semiotic) processes in human interactions with robots?

All papers collected in this special issues address both of these questions, albeit with different emphasis. To highlight the theoretical dimension of robophilosophy, in contrast to robo-ethics, we have ordered the contributions in a way that leads the reader from more conceptual-theoretical to more practical-normative investigations. However, it seemed productive to open the forum of debate with the contribution by Kerstin Fischer, who offers in "Why Collaborative Robots Must Be Social (and Even Emotional) Actors" what may be considered a 'proof of concept' for this special issue. Fischer, an internationally renowned Human-Robot Interaction researcher with longstanding experience, argues that as soon as we aim to engage with robots in any form of collaborative activity (as opposed to pre-programmed coordination with fixed adaptation on the human side, as with traditional industrial robots) we need to endow robots with relevant cues for social perception. Since "explicit and implicit modes of communication interact during successful and effortless cooperation," the call for transparency, i.e., for interac-

tion designs that do not dress up machines as 'social' agents, is both impractical and restricts access in ethically problematic ways: "a decision against the use of social signaling systems may not only make robots tiresome to use, but also prevent large groups of the population from interacting with them at all" (Fischer 2019, 276). As Fischer argues based on empirical evidence from linguistic and psychological discourse analysis as well as HRI, perceptual cues about emotions (i.e., about emotions experienced or perceived) carry important social information about agentive intentions and expectations. Since "social and emotional signals are human societies' shortcuts to solving complex coordination tasks at different levels simultaneously" (Fischer 2019, 281), and since the social and emotional dimensions of these cues cannot be easily separated, it seems that both for practical but also for theoretical reasons the 'problem of emotions' in artificial sociality cannot be circumvented.

Our descriptions of robots, their features and doings, either are outright metaphorical or exploit in problematic fashion extant ambiguities—between functional *versus* intention-presupposing understandings of 'what is done,' between physical descriptions of bodily aspects and movements and their semiotic significance, and even ambiguities between the material and the phenomenological senses of terms for parts of the human organism. What philosophers can contribute to clarify these problematic *quid pro quo*s comes out in the second and third contribution, both addressing the link between bodily expressivity and emotional affordances.

The term 'human skin' can be understood as the label for a certain cell structure or human organ that envelops most of the other cells constituting a human body. In that sense of 'skin,' i.e., as a certain material that functions as a boundary between two regions, robots can have a skin. However, as Van Grunsven and Van Wynsberghe point out in "A Semblance of Aliveness: How the Peculiar Embodiment of Sex Robots Will Matter," human skin supports a host of phenomenological experiences that are constitutive for our self-perception as beings in a world—"human skin is not merely a biological marker of separateness and exposure, but that it has a profoundly *social* and *existential* significance that gets at the heart of our ties to and vulnerabilities in the face of the other" (Van Grunsven and Van Wynsberghe 2019, 297). The perception of touched skin may be constitutive for the experience of empathy, and our feeling of social connectedness may be anchored in the reciprocity of our perceived vulnerabilities at the skin, the "precarious boundary between inner and outer" (ibid., 298) of our lived bodies. Combining the perspectives of embodied cognition research and second-person cognitive science with phenomenological observations, the authors argue that the

full experiential dimension of human sexual intimacy hinges on (i) the experience of 'skin' in this sense of an affordance for the reciprocal experience of vulnerability, (ii) on the rich scope and fine-grained expressivity of the human body, and (iii) the fact that our human-human interactions realize skills, not programs, and thus constitutively incorporate experiences with and expectations of spontaneous deviations. Since sex robots currently—and perhaps in principle—cannot fulfill these prerequisites, there is a clear risk that the use of sex robots will impoverish our skills for human sexual intimacy. While acknowledging this risk the authors suggest that we may also explore instead the possible benefits of sex-robots for people on the autism spectrum who have diminished capacities to engage in the "intercorporeal dialogues" (ibid., 301) in which emotions are bodily expressed.

While Van Grunsven and Van Wynsberghe identify the problematic ambiguities of the term 'skin' as used by robotics engineers and in illuminating fashion elaborate on the phenomenological role of touch and human skin, the contribution by Jaana Parviainen, Lina van Aerschot, Tuomo Särkikoski, Satu Pekkarinen, Helinä Melkas, and Lea Hennala, "Motions with Emotions? A Phenomenological Approach to Understanding the Simulated Aliveness of a Robot Body," focuses on the ambiguity of the term 'body.' In robotics engineering, and in particular "morphological computing" and "affective computing," the term "body" is used in descriptive contexts that relate material and kinematic aspects to certain functional goals—either instrumental goals, or communicative goals, such as simulating the expression of human emotions or eliciting human emotions. By contrast, in common-sensical references to the human "body" the term often denotes something else, the "lived body," an experiential unit for first-person and second-person experience. In phenomenological parlance following Husserl, the "lived body," in German "Leib" as opposed to "Körper," is an experiential unit that carries conscious and subconscious communicative functions in expressing and eliciting emotions. However, as the authors point out, following Helmuth Plessner, also the physical body (Körper) has semiotic functions in human social cognition and interaction (e.g., by providing perceptual cues about gender, age, ethnicity etc). While the two focal readings of "body"—physical body and lived body—should be distinguished, the authors argue, our social perceptions and practices are structured by both conceptions: we have a "double body." This "double body effect" can also be traced in the phenomena of human-robot interaction even though robots "lack lived bodies (*Leib*) as they do not have the capacity for bodily awareness" nor "have physical bodies (*Körper*) of the kind that phenomenologists define as living, biological and physiological entities" (Parviainen et al. 2019, 327). Robots are

"material things" as well as "simulations of the lived body," thus they also have a "hybrid robot body" (Parviainen et al. 2019, 319). Reporting on a field study with the robot ZORA, the authors trace in detail how the simulation of a lived body is largely generated by the narratives of the users of the robot to support the robot's performance but also suggest that the emotions engendered by the robot hinge on aspects of both the physical and the lived body—material aspects, the morphology, the simulation of animate movements, and on the user's narratives.

The fourth paper, "The Dilemma of Openness in Social Robots" by Felix Lo, demonstrates that the philosophy of social robotics does not need to stay close to the idioms and interpretatory frames of analytical philosophy or phenomenology in order to offer productive suggestions for the design of social robots. Leading the reader to the tradition of continental philosophy of culture and technology, Lo presents two accounts of the "openness" of technology, anchored in the speculative ontology of Gilbert Simondon and in Umberto Eco's reflections on art, respectively. Lo uses in particular Simondon's descriptions of the conditions for transindividual inter-human relations in order to explain why some types of social robots may promote personal and social growth while others do not. As Lo suggests, social robots with non-humanoid form such as PARO can generate an "emotional coupling" between human and robot that have the potential to engender human-human relationships via the robot. By contrast, once social robots are given humanoid form, these "the emotional intensities and the confrontations in human relationships [that according to Simondon] are the very source of potentiality for personal and social growth" do not arise since the perceptual conflict of a machine in human-like form is emotionally disturbing, uncanny (Lo 2019, 359). Lo suggests that the 'openness' or productive potential of social robots thus depends on choosing forms and functionalities that enable positive emotions—especially by avoiding close resemblances to the human form. The paper also contains, more indirectly, the further suggestion that social robots that are designed in ways that fit with the criteria of contemporary art, as offering only "incomplete decoding schemes" (Lo 2019, 356) and thus require the active engagement of the interpreter, will more likely engender positive emotions with transformative potential.

Ethical discussions about social robots are hampered by the fact that currently they cannot be based on the kind of empirical research that would be necessary to give concrete recommendations rather than suggestions based on hypothetical scenarios (Rodogno 2017). Since social robots are not yet entrenched widely within our societies, there are no sufficiently reliable empirical data about longterm effects on human well-being, nor on the intrapersonal variation of such

effects. This state of ignorance is connected to what in the technological design community is known as the "Collingridge Dilemma"—we are about to release into society a technology whose consequences can only appraised at a time when the technology cannot be extracted anymore (Collingridge 1980). Precisely due to the link between embodiment and emotion it is plausible to think that social robotics technology presents an even better illustration of the Collingridge Dilemma than smart phones and social media in the sense that this technology will be even less inextractable once broadly entrenched. But also the other element of the Collingridge Dilemma, the incalculability, arises here in a particularly exacerbated form, given that we do not even have suitable descriptive categories for 'what is happening in human-robot interaction' here and now. In order to react to this exacerbated variety of the dilemma, the "triple gridlock of description, evaluation, and regulation" in social robotics, several measures are necessary in which theoretical and practical philosophy has an important role to play (Seibt 2016; Seibt, Damholdt, and Vestergaard 2018). One of the philosopher's tasks is to pursue a double strategy in the ethical discussion about social robots. The ethical debate must either focus on particular cases, focusing on the available empirical results in HRI; this approach, let's call it 'strategy 1,' must be undertaken with keen attention to the fact that the 'so-called ethical issues' observed by robotics engineers—who often equate 'ethical issues' with 'safety issues'—but also of psychologists, anthropologists and other researchers involved in HRI and social robotics may differ from the kinds of concerns a professional ethicist would raise. Alternatively, or in addition, the ethical discussion must explore our (i.e., a certain community's) ethical imagination and normative sensitivities via thought experiments and relevant comparisons; let's call this 'strategy 2.'

Among the many ethical questions discussed in robo-ethics, one issue has been standing out so far—the problem of dignity. The interaction with social robots could negatively affect human dignity, it has been claimed. The very idea of *social* robots seems to violate the traditional reciprocity conditions of social interactions and involves people in a normatively 'deviant' activity that is either degrading in itself or since it is engendered on the basis of deception. That asymmetric sociality (Seibt 2017b) can be in itself degrading is most easily argued with reference to inappropriate emotional attachments. While in some contexts our emotional relations to social robots could be construed as innocuous sentimental attachments—as we have them to other inanimate items such as a city, landscape, thing, piece of music, or fictional characters (Rodogno 2016)—in other contexts the attachments and emotional dependencies evoked by robots are directly tied up

with presuppositions of reciprocity of emotion or commitment and thus harbor emotional harm even when no deception is involved.

The three ethical papers included in this issue address the topic of dignity and its relation to emotions in interestingly different ways. Arto Laitinen, Marketta Niemelä, and Jari Pirhonen use what we called strategy 1 above and offer for a special application domain, relative to the empirical research available for this domain, a detailed and conceptually precise discussion of the question of whether the use of robots used in the care practices of elderly care necessarily should be understood as a threat to dignity. The authors point out that "nothing can challenge human dignity as an inalienable status and as a source of demands" (Laitinen, Niemelä, and Pirhonen 2019, 376). Our sense of dignity, however—which they consider to be an emotion, as well as other emotions that our sense of dignity is connected with, e.g., shame, pride, joy, sadness, anger—can be affected by the way in which others *de facto* recognize our dignity, i.e., realize the duties that follow from the demands of dignity. Distinguishing among three modes in which robots can be used in care practices ("robot-based interaction in care, robot-assisted interaction in care, and teleoperated interaction in care") the authors investigate, in detail and in close contact to relevant empirical research in HRI, how these three modes of using 'care robots' relate to three dimensions of the recognition of dignity—the recognition of a person's vulnerability, their sense of agency, and their capacities as experiential subjects, i.e., as feeling and thinking subjects. The upshot of these investigations is a highly differentiated evaluation with a host of constructive suggestions.

The contribution by Sven Nyholm and Lily Eva Frank (Nyholm and Frank 2019) also addresses the issue of dignity and inappropriate emotional attachments also with focus on a particular application context, namely "sex robots." However, given the lack of empirical research, the authors use what we called strategy 2, a largely speculative approach. The authors focus on what appears to be a particularly objectionable feature of current design goals in sex robotics, namely, the fact that sex robots are intentionally designed and explicitly canvassed as romantic partners that express and understand emotions. The authors discuss three scenarios in which ethical worries about sex robots might arise because of the expressed design goal to create affordances for human emotional attachments. The design goal of creating such emotional attachments could be ethically problematic, the authors argue, if these attachments are used for financial exploitation, involve deception, or prevent the user from having more fulfilling sexual relationships with other humans. In a second step the authors also consider which, if any, recommenda-

tions could be given to the developers of sex robots to make their products more acceptable in ethical regards.

The third robo-ethical contribution also pursues a largely speculative-hypothetical approach exploring our ethical imagination, but here we are asked to imaginatively explore the special aspects of human-robot interactions as a possible source of enhancement of our moral dignity. In "Robot Nudges and Moral Improvement through Stoic Practice," Michał Klincewicz invites us to consider the positive potential of the peculiar way in which our cognitive 'mechanisms' are affected when we interact with robots. As researchers in neuropsychology and cognitive science have begun to explore in greater detail, social robots trigger pre-conscious 'mechanisms' and pre-conscious but consciously accessible 'routines' that are involved in human social cognition (Wykowska, Chaminade, and Cheng 2016). This phenomenon carries a potential for manipulation which the social robotics and HRI community has begun to discuss under the label of "robot nudging." Robot nudging exploits the cognitive mechanism of human imitation learning but it can also be combined with interventions for active reflection to channel the manipulatory elements into a pedagogical process. It is this explicitly pedagogical understanding of 'robot nudging' that Klincewicz has mind, human-robot interactions that incline towards "conscious reflection [of certain options] without compromising their target's ability to choose to do otherwise" (Klincewicz 2019, 426). Klincewicz critically discusses recent proposals for moral improvement by robot nudging that aim to dispose towards particular beliefs and attitudes in moral agents. Instead, Klincewicz argues, robot nudging should proceed in a more "theory-neutral" fashion and support the formation of the more fundamental "capacities relevant to moral behavior and moral decision-making" as such (Klincewicz 2019, 435). Empirical studies in developmental (neuro-)psychology can provide important insights into the relevant capacities and mechanisms of moral cognition—e.g., mechanisms for affective perspective taking—that a robot nudger could support, Klincewicz argues, but the identification of what is to count as 'moral behavior' and 'moral decision making' as used in these empirical disciplines is itself open for discussion. Here Klincewicz recommends that we revisit the "practices" of Stoic philosophy, which are also taken up in cognitive-behavioral therapy, as target dispositions that could be acquired by way of robot nudging. Interestingly, as Klincewicz elaborates, Stoic practices guide towards a particular emotional state or attitude, a frame of mind characterized by concern for a wide scope of others.

In the last section of his paper, Klincewicz discusses whether social robots are really needed to facilitate moral improvement and lists a number of aspects

that are idiosyncratic to human-robot interaction. Similar considerations arise in most of the other contributions—there is something that robots can do that humans can't. (As one of us has argued, design and development of social robotics applications should abide by a "non-replacement principle: robots may only do what humans should but cannot do" (Seibt, Damholdt, and Vestergaard 2018, 37). But the insight that robots hold special potential holds not only with respect to future practices but also in theoretical and even in metaphilosophical regards. Remarkably, artificial agents that—as of yet: merely—simulate social agency and linguistic competence without the accompanying phenomenology, without conscious experience, without emotions, can reconfigure our inquiry into the role of emotions in social interactions with embodied agents.

References

Breazeal, Cynthia. 2002. *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Collingridge, David. 1980. *The Social Control of Technology*. London: St. Martin's Press.

Fellous, Jean-Marc, and Michael A. Arbib, eds. 2005. *Who Needs Emotions?: The Brain Meets the Robot*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195166194.001.0001

Fischer, Kerstin. 2019. "Why Collaborative Robots Must Be Social (and Even Emotional) Actors." *Techné: Research in Philosophy and Technology* 23(3): 270–89. https://doi.org/10.5840/techne20191120104

Gilbert, Margaret. 1992. *On Social Facts*. Princeton, NJ: Princeton University Press.

Jerčić, Petar, Wei Wen, Johan Hagelbäck, and Veronica Sundstedt. 2018. "The Effect of Emotions and Social Behavior on Performance in a Collaborative Serious Game between Humans and Autonomous Robots." *International Journal of Social Robotics* 10(1): 115–29. https://doi.org/10.1007/s12369-017-0437-4

Klincewicz, Michał. 2019. "Robotic Nudges for Moral Improvement through Stoic Practice." *Techné: Research in Philosophy and Technology* 23(3): 425–55. https://doi.org/10.5840/techne2019122109

Kolling, Thorsten, Stefanie Baisch, Arthur Schall, Stefanie Selic, Saskia Rühl, Ziyon Kim, Holger Rossberg, Barbara Klein, Johannes Pantel, and Frank Oswald. 2016. "What Is Emotional About Emotional Robotics?" In *Emotions, Technology, and Health*, ed. Sharon Tettegah and Yolanda Garcia, 85–103. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-12-801737-1.00005-6

Laitinen, Arto, Marketta Niemelä, and Jari Pirhonen. 2019. "Demands of Dignity in Robotic Care: Recognizing Vulnerability, Agency, and Subjectivity in Robot-

Based, Robot-Assisted, and Teleoperated Elderly Care." *Techné: Research in Philosophy and Technology* 23(3): 366–401.
https://doi.org/10.5840/techne20191127108

Lisetti, Christine, and Eva Hudlicka. 2015. "Why and how to build emotion-based agent architectures." In *The Oxford Handbook of Affective Computing*, ed. Rafael Calvo, Sidney D'Mello, Jonathan Gratch and Arvid Kappas, 94-109. New York: Oxford University Press.

Lo, Felix Tun Han. 2019. "The Dilemma of Openness in Social Robots." *Techné: Research in Philosophy and Technology* 23(3): 342–65.
https://doi.org/10.5840/techne20191126107

Miller, Seumas. 2001. *Social action: A Teleological Account*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511612954

Novikova, Jekaterina and Leon Watts. 2014. "A Design Model of Emotional Body Expressions in Non-Humanoid Robots." HAI '14 *Proceedings of the second international conference on Human-agent interaction* 353–60. New York: ACM
https://doi.org/10.1145/2658861.2658892

Nyholm, Sven, and Lily Eva Frank. 2019. "It Loves Me, It Loves Me Not: Is It Morally Problematic to Design Sex Robots that Appear to Love Their Owners?" *Techné: Research in Philosophy and Technology* 23(3): 402–24.
https://doi.org/10.5840/techne2019122110

Parviainen, Jaana, Lina van Aerschot, Tuomo Särkikoski, Satu Pekkarinen, Helinä Melkas, and Lea Hennala. 2019. "Motions with Emotions? A Phenomenological Approach to Understanding the Simulated Aliveness of a Robot Body." *Techné: Research in Philosophy and Technology* 23(3): 318–41.
https://doi.org/10.5840/techne20191126106

Rodogno, Raffaele. 2016. "Social Robots, Fiction, and Sentimentality." *Ethics and Information Technology* 18(4): 257–68. https://doi.org/10.1007/s10676-015-9371-z

Rodogno, Raffaele. 2017. "Robots and the Limits of Morality." In *Social Robots*, ed. Marco Nørskov, 39–55. New York: Routledge.
https://doi.org/10.4324/9781315563084-3

Saldien, Jelle, Kristof Goris, Bram Vanderborght, Johan Vanderfaeillie, and Dirk Lefeber. 2010. "Expressing Emotions with the Social Robot Probo." *International Journal of Social Robotics* 2(4): 377–89.
https://doi.org/10.4324/9781315563084-3

Scheutz, Matthias. 2014. "Artificial Emotions and Machine Consciousness." In *The Cambridge Handbook of Artificial Intelligence*, ed. Keith Frankish, and William M. Ramsey, 247–66. Cambridge: Cambridge Univ Press.
https://doi.org/10.1017/CBO9781139046855.016

Searle, John. 2009. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:osobl/9780195396171.001.0001

Seibt, Johanna. 2016. "Integrative Social Robotics—A New Method Paradigm to Solve the Description Problem and the Regulation Problem?" In *What Social Robots Can and Should Do—Proceedings of Robophilosophy/TRANSOR 2016*, ed. Johanna Seibt, Marco Nørskov and Søren Schack Andersen, 104–14. Amsterdam: IOS Press.

Seibt, Johanna. 2017a. "Robophilosophy." In *Posthuman Glossary*, ed. Rosi Braidotti and Maria Hlavajova, 390–94. London: Bloomsbury.

Seibt, Johanna. 2017b. "Towards an Ontology of Simulated Social Interaction: Varieties of the 'As If' for Robots and Humans." In *Sociality and Normativity for Robots*, ed. Raul Hakli and Johanna Seibt, 11–39. New York: Springer.
https://doi.org/10.1007/978-3-319-53133-5_2

Seibt, Johanna, Malene F. Damholdt, and Christina Vestergaard. 2018. "Five Principles of Integrative Social Robotics." In *Envisioning Robots in Society: Power, Politics, and Public Space*, ed. Mark Coeckelbergh, Michael Funk, Janina Loh, Johanna Seibt, Marco Nørskov, 28–43. Amsterdam: IOS.

Tuomela, Raimo. 1984. *A Theory of Social Action*. Synthese Library vol 171. Dordrecht: Reidel. https://doi.org/10.1007/978-94-009-6317-7

Vallverdú, Jordi. 2009. *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*. Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-60566-354-8

van Grunsven, Janna, and Aimee van Wynsberghe. 2019. "A Semblance of Aliveness: How the Peculiar Embodiment of Sex Robots Will Matter." *Techné: Research in Philosophy and Technology* 23(3): 290–317.
https://doi.org/10.5840/techne20191125105

Wykowska, Agnieszka, Thierry Chaminade, and Gordon Cheng. 2016. "Embodied Artificial Agents for Understanding Human Social Cognition." *Philosophical Transactions of the Royal Society B* 371(1693): 20150375.
https://doi.org/10.1098/rstb.2015.0375