Alienation and Recognition: The Δ Phenomenology of Human-Social Robot Interactions (HSRI)

Piercosma Bisconti[1] & Antonio Carnevale[2]

[1] Sant'Anna School of Advanced Studies, Pisa, piercosma.biscontilucidi@santannapisa.it
[2] CyberethicsLab, Rome, a.carnevale@cyberethicslab.com

Table of contents

Abstract

A crucial philosophical problem of social robots is how much they perform a kind of sociality in interacting with humans. Scholarship diverges between those who sustain that humans and social robots cannot by default have social interactions and those who argue about the possibility of an

Both authors contributed equally to this paper.

asymmetric sociality. Against this dichotomy, we argue in this paper about a holistic approach called "Δ phenomenology" of HSRI (Human-Social-Robot-Interaction).

In the first part of the paper we will analyse the semantic of a HSRI, that is what leads a human being (*x*) to assign or receive a meaning of sociality (*z*) by interacting with a social robot (*y*). Hence, we will question the ontological structure underlying HSRIs, suggesting that HSRIs may lead to a peculiar kind of users alienation. By combining all these variables, we will formulate some final recommendations for an ethics of Social Robots.

Keywords

Companion Robots, Human-Robot Interactions, Robo-Ontology, Value-Based Design, Robo-Ethics, Alienation, Recognition,

Acronyms and abbreviations

Acronym	Title
HRI	Human – Robot Interaction
HSRI	Human – Social Robot Interaction
HHI	Human – Human Interaction
SR	Social Robot
CR	Companion Robot

1. Introduction

Although the concept of robot is a constantly moving target that constrains us to reinvent what we consider to be one (Dautenhahn 2014; Siciliano & Khatib 2016), many scholars have begun to define what could or should be understood for qualifying as "social" an interaction between a human and a robot. Theoretically speaking, this relationship should be imagined as an enabling communication that allows people to interact with the machine "as if it were a person, and ultimately as a friend" (Breazeal 2002). Clearly, this kind of assumption has raised a large philosophical debate so far.

Scholars are preoccupied about ethical and moral aspects, either sidestepping conceptual issues or addressing them obliquely within a specific context of application (Sharkey & Sharkey 2012; Sparrow & Sparrow 2006; Coeckelbergh 2010; Vallor 2011) and supporting the ethical consideration by empirical research on human-robot interaction (Kahn et al. 2004). In our view, before advancing any sort of ethical assessment or normative judgement, a philosophical investigation should commence by inquiring the phenomenology of social robotics with a basic question: What do we semantically refer to when we imagine or think to HSRI?

2. The Debate: Referring to HSRI "as" an experience of sociality

On the interaction level, understanding how users operate the semantic attribution of sociality is crucial: we can grasp, from this attribution, a user's pre-disposition towards the robot, as also Coeckelbergh (2011) points out when discussing the role of pronouns in HSRI. Any possible remarks should not ignore a fundamental pre-consideration: the asymmetry that invests any HSRI by default makes it intrinsically difficult to attribute "sociality" to any interaction between a machine and a human being. Seibt (2017) has identified different pragmatic and semantic models by which a possible HSRI can be shaped. They can be summarized by three different classificatory attributes of HSRI in which (1) "*x treats y as z*", or (2) "*x interacts with y as if it were z*", or (3) "*x takes y as z*". In the next paragraphs we illustrate the three semantic attribution models designed by Seibt, a starting point for the subsequent discussion on HSRI.

The first attribute is *make-believe* and it refers to a type of relationship centred on human perception and habits. *X*, the human, *treats Y*, the robot, *as if it were someone else*, *Z*. The human makes inferences of sociality in HSRI since he/she believes and accepts the socially mimetic behaviour of the robot. The interaction is more the result of a mimesis or, more precisely, of an interpretative capacity of *X* who, importing into HSRI emotional contents coming from outside, mostly from her/his own personal story, ends up associating or projecting meanings of otherness to *Y*. In doing so, the

artificiality of the sociality experienced in HSRI is dissimulated. As we will point out later, in this kind of semantic attribution the phenomenology of HSRI appears in the form of an alienated sociality.

The second attribution of meaning is affected by the power of the *fiction* and *gameplay*. Here the meanings of sociality are performed within the interaction, that is, they do not come from external factors. This time, such a performativity takes more into consideration the existing distinction between a social relationship believed to be such ("*x treats y as z*") and the attribution of simulation as an embedded cognitive value of the HSRI ("*x interacts with y as if it were z*"). In other words, *X* and *Y* can interact only to the extent that the fiction that holds the relational scaffolding is accepted as *a rule of the game*. This is the tacit awareness of this fictional condition that allows people to simulate a social interaction, and so play *as if* they can have interactions with the robot. Like most games, this means that the rules are set before the play begins.

The third model for attributing social significance to HSRI is the one designed for the *socially instituted interaction*. The human, *X*, approaches the robot, *Y*, making previous assumptions and in accordance with extant social conventions, so that *X* can take *Y* no longer as *something else*, as in the first case, but as *something*, that something that defines a robot as a robot. In accordance with this assumption, the semantic attribution of sociality is not solely performed in terms of fiction ("*x interacts with y as if it were z*"), but it is performed by a constructivist phenomenology of the HSRI or, in our terms, that humans become aware that ***social robotics is a socio-technical system and consists in parts of a sharing discourse that needs requirements and rules to be decided for having social exchange*** in HSRI. Highlights of (prosing) settings of rules will be presented in the last part of this paper.

Several current philosophical analyses of social robotics mover away from the notion that there are several competing notions of sociality and just a singular one. Thus, "we should abandon the idea of a dualist distinction between social and non-social interactions; rather, we should conceive of sociality as a matter of degree" (Seibt 2017: 15). Similar accounts are undeniable. Nevertheless, on

the background of the argumentation held in this paper, further considerations move our stance to reframe the point. The degree of sociality in HSRI is often understood as a classificatory cohort of semantically different occurrences. In other words, HSRI are imagined as a division of types of ideally-simulated experience that occur in a sort of Cartesian taxonomical HSRI space, organized around a *Rex extensa* (the semantic type of interaction simulated by the robot) and a *Rex cogitans* (the type of social value to attribute to such an experience). In so doing, however, the classificatory attempt, because of its lack of conceptual foundation, drives the philosophical analysis to make a deductive use of the "ontological argument" – in the same way that Descartes derived the ontological proof of the existence of God from the conceptual division of the things and the thoughts. Not surprisingly, the philosophical debate on social robotics ended up dividing between "hard" and "soft problem" of ontology of simulated HSRI. Indeed, the ontology of HSRI can enlighten us both on the hard question of what, if anything, is lost "when a process is perfectly simulated"; more softly, the ontology of HSRI could refer to how the discussion of taxonomic questions "might help us to decide whether and where the hard problem matters at all" (Seibt 2017).

3. Our philosophical background: HSRI as a Δ phenomenology

We will discuss the ontological question later. For the moment we want to make a preliminary and preparatory consideration. The hard-soft problem of ontology induces a choice between two options. On one hand, since current philosophical definitions of social interactions determine that, for a social interaction to occur, all agents involved must have the capacities required for normative agency, e.g., intentionality, consciousness, normative understanding (Hakli 2014), we should conclude that HSRIs are by default non-social interactions. Alternatively, on the other hand, we face the challenge and choice of building new conceptual tools for forms of non-reciprocal or asymmetric sociality, i.e., "for social interactions where one agent lacks the capacities required for normative agency" (Seibt 2017: 11).

Against this dichotomous perspective, we argue that the philosophical question about HSRI is neither in rejecting *tout court* the hypothesis of emerging sociality in the relationship between humans and robots, nor in inventing new standards of sociality that adapt human life to HSRIs. Rather, the question is returning these two interpretations in the frame of different options of a philosophically constructive discourse on the self-production of attributable meanings within socio-technical systems, capable of recognizing the breaking point in which a *consciously and task-related simulation of sociality* is transformed into an *alienated form of phantasmatic human-machine intersubjectivity*. We do not see the denial of sociality, on one side, and the creation of a new sociality, on the other side, as two opposite poles of understanding HSRI. *Their opposition is not the solution, rather the question.* To this question, we oppose in this paper a solution of a Δ phenomenological type. Δ phenomenology of HSRI is the differential between all possible forms of *human-centred* [2] sociality that can be set up between a human and a social robot – regardless of the fact that the sociality might be a mystification of the user or a function specifically designed in the interface of a robot. Otherwise stated, Δ phenomenology wants to approach the consistency of ontology (hard vs. soft) not as a presupposition of the philosophical investigation, rather as its own component. Instead of beginning from a semantics that elicits separate logical occurrences of sociality in HSRI, in our view, Δ phenomenology needs to deal with HSRI in terms of the phenomenology of a self-posing socio-technical system. This system should be not presumed as a whole, rather as a ***sharing discourse that constructively draws off sense and meaning from the different forms of experience in which HSRI can be realized by simulation, banally from the coherent-by-design interaction with a robotic social machine to the unfulfilled desire for recognition that the subject places on the symbolic and psychological value of the interaction with the machine.***

[2] The human-centred perspective does not want to be a value-based assumption. It is not for a *petitio principii* that we adhere to this perspective, nor for a trivial exercise of the precautionary principle (it is better to always favour forms of sociality that give priority to the human being). What moves us at the base is a principle of "*technological realism*": to date there are no robotic forms that can suggest the danger of the onset of an anti-human sociality.

In this wake it also insists the meaning we give to the term "alienation". Certainly, its philosophical connotation is inspired by the twentieth-century Hegelian-Marxist and existentialist matrix (Wood 2004) which links the idea of alienation to that of a loss and separation from the 'true self', a cause of estrangement, a distortion of the human essence, a loss that, therefore, implies the necessary re-appropriation of what has been lost. This approach, however, inevitably leads to the adoption of an objectivist and substantialist position, in some sense fetishist – consider the example of Christian religious consciousness, as broadly understood in the writings of Ludwig Feuerbach. To this connotation, we integrate a definition of non-substantial but formal alienation. In the wake of the studies of Rahel Jaeggi (2014) and Axel Honneth (2007), the alienation to which HSRIs can lead does not designate a specific content, but a disturbed relationship with oneself and with the world. More precisely: alienation is an obstacle to the possibility of being able to dispose of oneself and one's world. In other words, ***alienation is not losing one's human nature, rather the freedom to be able to dispose of it***.

From the conventional interaction until the alienation, the logical particle "as" ("x treats y *as* z", "x interacts with y *as* if it were z", "x takes y *as* z") no longer denotes the basic semantic infrastructure for which attributing sociality is never practically possible. Rather the particle "us" recalls, in our opinion, the formulation that Heidegger has given in his well-known analysis of what he calls the *Als-Struktur* (Ricœur 1968). Heidegger distinguishes two nomenclatorial significances of *Als*: the existential-hermeneutic, which is primordial and concerns the projection of interpretation onto the determining horizon of language; and the apophantic-ontological, derivative of the latter, which concerns the determination of the human being in the mode of truth's determination. With the first sense of "as" we translate the thing in our interpretative horizon, with the second sense we examine whether our interpretation, according to our own horizon of interpretation, matches the given conceptual – in case of social robots, socio-technical – frameworks or not. Likewise, "as" opens construction plans of variable meanings, which in this article we have tried to formulate as follows:

- ***x = the human being who interacts with the social robot***
- ***y= the social robot designed to respond coherently by simulating some degree of sociality***
- ***z = the meaning of social interaction that can assume a certain ratio of x → y* (the differential between x and y)**

And, given these variables, we define:

- ***Δ (y, z) = the phenomenological differential between the coherent-by-design sociality simulated by the robot and the meaning of sociality that x receives or attributes by acting with y***.

Before moving further in our discussion, a clarification must be made here. Evidently the least understandable element of the previous formulation is the value of "z". Who is performing the value of z? The robot? Certainly not, because, apart from any reasonable philosophical doubt about predicting "sentient" robots (Haikonen 2013), there is no current example of a robot that can give us a semantic account of the social value that it attributes to some operation with a human being.

If it is not the robot, could it then be the human experiencing the HSRI? Yes, but from what perspective? The subjective and internal standpoint of the experiencing individual? For psychologists this perspective would certainly be the best. However, from our perspective it does not help, since the subject could experience sociality in HSRI because subjectively altered by external factors to the experience itself. It is true that such an outcome also happens to non-HSRI experience. How many of us have experienced wondering whether a Platonic love for another human was reciprocated? However, during HHI we know of the possibility of internal reflection, even if subjectively given, and that we are in a position to mitigate the alteration of the sensitivity of the experience and restore realism to it. With robots we have excluded this faculty at the moment since, not it has not yet been technologically designed in the machines. In fact, in HHI one's semantic attributions to the relation are limited to those that the other subject is willing to accept. In other words, there occurs a semantic

“bargain” over the relationship between the two subjects. In HSRI, the robot does not limit its user’s semantic attribution, if not by objective limitations.

To understand the value of “z” we should look for a type of assignment of meanings outside the subjectivism of personal experience. Therefore, it would be better to abandon the perspective of the first person and opt for that of the second and third person who appear to be based on a more shared semantically stable perspective. In fact, for being a “you” and a “he” (or a “she”), the assertion should transit from the simple plane of meaning attribution (“The robot and I understand each other”) to the sociolinguistic plane in which the same attribution is recognized and assumed as a cognitive commitment by a participant in front of an audience ("You know what? The robot and I understand each other!”). This happens because when we communicate, we externalize our intentions (Grice 1957). Language is not a simple denoting of things with words, but a moral performance that must be acted out to reach a sharing of meanings that we can call “semantic inferentialism” (Brandom 2000). And it is precisely because we foresee that a semantic attribute has the social dimensions of language incorporated within it, that the logical particle “as” can now be moved and represent different kind of phenomenologist attribution to *“z”*: from a meaning of maximal possible social interaction – “*x treats y as z*” – to a meaning of minimal ones – “*x takes y as z*”.

As we have shown, z occurs based on x’s stated intention of interacting with y in a given way. We have offered more clarifications on the value of “z” which, as we have seen, at the same time also informed us on the type of experience of “x”. What remains to be done is to close the circle and widely clarify the value of “y”. y is the social robot that we settle according to the definition of the *Handbook of robotics:*

“Social robots are designed to interact with people in a natural, interpersonal manner – often to achieve positive outcomes in diverse applications such as education, health, quality of life, entertainment, communication, and tasks requiring collaborative teamwork. The long-term goal of

creating social robots that are competent and capable partners for people is quite a challenging task. They will need to be able to communicate naturally with people using both verbal and nonverbal signals. They will need to engage us not only on a cognitive level, but on an emotional level as well in order to provide effective social and task-related support to people. They will need a wide range of social-cognitive skills and a theory of other minds to understand human behaviour, and to be intuitively understood by people.*"*

(Breazeal, Dautenhahn, & Kanda 2016: 1935)

First block of considerations

Following these assumptions, it can be argued that:

- The lower this differential is[3], the more the robot will perform a simulated task-related sociality that is coherent by-design.
- The consequent form of sociality is made up of a type of interaction in which *x* acts with *y* assuming it for an interactive technological artifact in the world.
- On the contrary, the value of the differential expands when the difference between the coherent-by-design sociality simulated by the robot and the meaning of (derived or attributed) sociality of *x* interacting with *y* becomes an additional experience level beyond the task-related ones defined by design.
- The consequent form of sociality is made up of a type of interaction in which *x* acts with *y* assuming it for an interactive techno-entity possessing some degree of otherness.

4. From the reference to the experience of HSRI

The realm of sociality of HSRI basically consists in adapting to discursive practices and norms in the sphere of social interaction. The systemic nature of socio-technical systems implies that any actor

[3] That is, "z" satisfies the social functionalities for which *y* is coherently designed.

who performs context-coherent behaviours makes significant changes to the system. Therefore, if we define sociality solely as a co-modification between actors in a context with interactional pragmatic rules, SR can be defined as social. They can be said to be less or more social, as we saw with the three possible uses of "as" in interacting with a robot. In each of these cases, their presence in the interactional space produces changes in the system, modifying social practices and the balance of the actors' interactional network. Social interactions can be a powerful mode through which a robot performs a useful action: Siri interacts in order to give information considered by the user to be practical; robotic waiters (Asif et al. 2011) interact with users with the goal of waiting tables where they are. Nevertheless, social interactions are not simply instrumental, but imply emotional bonding and forms of attachment. Here we mark a conceptual shift from interacting to relating, where intersubjectivity is implied (Cassel & Tartaro 2007). In the previous chapter we discussed the criteria of attributing coherent-by-design sociality to HSRI, in this chapter we enquire the possibility of attributing intersubjectivity to HSRI. Under this lens, we take as a striking example a specific subclass of SRs, specifically designed to produce a relation with the user: companion robots (CRs). The main goal of these machines is the care of children, elderly, or people with disabilities. These machines would be not only actors and modifiers of an interactive system, but, philosophically, an alterity which a human subject can relate to and face (Coeckelbergh 2016). In fact, CRs should use their interactional skills not only to achieve a specific goal (ex: giving information on the best route to follow is the goal of the navigator, when verbally interacting), but also to build a relationship with the user. Therefore, we can say that while SRs produce interactions to simulate sociality, CRs have the aim of building intersubjective relationships[4], reproducing – and to some extent substituting (Turkle et al. 2006) – human ones. Therefore, the theoretical shift is from interactions to relationships: this is the second level of the attribution problem. Is it legitimate to attribute the status of an intersubjective relationship to the human-robot interaction, and under what conditions? Can a robot

[4] Obviously interacting and relating are not mutually exclusive: we only make a difference on the scopes of SRs (of which CRs are a subset) and we differentiate the angle of our theoretical inquiry.

produce a relationship, as well as an interaction? We call this the ontological problem of SRs, as it does not refer to the performative and third-person view problems of HSRI, but addresses what Seibt (2017) calls "the hard problem of robo-ontology, namely, the question of what, if anything, is lost when a process is perfectly simulated". In this chapter we try to respond to this problem under the theoretical framework of the theory of intersubjectivity and recognition.

In the vast and inexhaustible debate on the meaning of "intersubjective", a fixed point of a long and shared philosophical tradition, starting from Hegel, is the concept of "mutual recognition". In the encounter-clash between the two self-consciousness the subjective nature of the Other is revealed. What emerges in the experience of the subjects is the irreducible alterity of the Other: his/her structural autonomy (Hegel 1807). The conceptual category of recognition has been declined in many ways. In Hegel, the recognition implies the struggle for life or death, which produces the master-slave relationship that will eventually undergo the dialectical inversion. In Alexandre Kojève mutual recognition is understood as the desire for the other's desire, namely being desired by the Other (Kojeve 1980). More contemporary reflections on the category of recognition are made by Axel Honneth, who understands it as genealogically and psychologically distinct stages along which individual persons gain self-confidence, self-respect and self-esteem (Honneth 2004). Throughout the reflection on this concept, two elements of continuity emerge. First of all, the intersubjective relationship is characterized by a constitutive otherness that existentially questions the position of the subject: otherness alters those who experience it. Secondly, the mutuality of recognition is essential to produce an intersubjective relationship.

From this standpoint of analysis, we will proceed to enquire the structure of the HSRIs to understand if, under the category of recognition, there can be human-robot intersubjective relations. Accordingly, to understand the ontological problem of HSRI, we have to break the two elements highlighted in the category of recognition, namely alterity and mutuality. Therefore (1) can robots be experienced as an

"Other"? (2) Is the mutual recognition of this otherness possible? Namely, can a robot recognize my otherness?

Don Ihde was among the forerunners to investigate the problem of the relational artifact's "otherness" (Ihde 1990). At the heart of Ihde's post-phenomenological approach to technology is an analysis of various types of relations between human beings, technologies, and the world. Ihde investigated in which ways technologies play a role in human-world relations, ranging from being 'embodied' and being 'read', to being 'interacted with' and being in the 'background'. Apart from "background relations", in which technologies are the context for human experiences and actions (i.e. air conditioners and fridges), the other three relationships of humans and technology refer to SRs. In embodied relations technology becomes a part of our perceived body (as glasses) when we interact with the external world; in hermeneutic relations artefacts allow us to access information about the world otherwise unavailable, as for the microscope. On the other hand, quasi other relations do not necessarily involve the world; the artefact appears outside of its instrumental and coherent-by-design function:

What the quasi otherness of alterity relations does show is that humans may relate positively or presententially to technologies […]. Technologies emerge as focal entities that may receive the multiple attentions humans give the different forms of the other. In alterity relations there may be, but need not be, a relation through the technology to the world […]. The world, in this case, may remain context and background, and the technology may emerge as the foreground and focal quasi-other with which I momentarily engage.

(Ihde 1990, pg. 107)

Ihde's quasi-other is therefore no longer instrumental, as in embodiment or hermeneutic relations where technology mediates user relation with the world, but it appears "as other to which I relate". Is this otherness, the one Ihde speaks of, the kind of otherness we refer to in intersubjective

relationships? Certainly, when escaping the instrumental categories of embodiment and hermeneutic, the object can appear (*φαίνεσθαι* as Heidegger would say) in its otherness. On the other hand, we cannot trace the emergence of the intersubjective relationship within this framework: the object, even when it appears to the subject in its structural otherness, even if in the form of a total alien (Coeckelbergh 2016), remains inert and does not in turn recognize the human subject (Ramey 2005).

Starting from Ihde, David Gunkel and Mark Coeckelbergh deal with the problem of robot otherness (Gunkel 2007; Gunkel 2018; Coeckelbergh 2011; Coeckelbergh 2020). They investigate the category of machine otherness starting from a critical appropriation of Levinas. In fact, the otherness of the other, when it is experienced, is immediately "domesticated" in the categories of similar-different to me, becoming normality and sameness. How to preserve the encounter with otherness, if it is no longer alien to us? Following a Heideggerian suggestion, the being-present of otherness vanishes when the machine is used: when the machine is present-at-hand it can "*φαίνεσθαι*" ("manifest itself"); when it is ready-to-hand it can only be instrumental, *Gestell*. The being-present of the otherness is threatened by the reduction of the "other" to familiar gnoseological category (Gunkel 2016). However, as Coeckelbergh himself points out, the problem of the *violence* that reduces – domesticates – the experience of the otherness, pertains also to human relationships. Therefore, this issue concerns the concept of otherness in general rather than the specific otherness of the robotic artifact. David Gunkel (2016) well specifies the process of epistemic domestication: he underlines that we experience otherness inside the gnoseological categories we already know. This problem raises the issue of what Gunkel and Coeckelbergh call "gnoseological violence" : I am hindered in experiencing otherness because I pre-determine the actions and the determination of the other on the basis of my epistemic preconception, therefore alterity, may be lost in the process. According to these scholars, the otherness of the other acquires the value of sharable meaning only inside the subject's pre-structured framework of experience and therefore every time I experience otherness I am also "domesticating" it, ultimately reducing the "*alter*" to the "familiar" (Coeckelbergh 2016).

Nevertheless, in our view, the possible ways of human and social machine "encounter" are not solely predetermined by the subject's epistemic anthropo-normalized predetermination, but also by the anthropocentric design that builds SRs for the purposes of human being. The design of SRs is geared in providing the most familiar interaction possible, also in the non-verbal parts of an interaction (Hegel et al. 2011; Mumm & Mutlu, 2011; Mutlu et al. 2009). Therefore, not only the qualities of otherness are predetermined by the subjects (2016), but SRs are predetermined by design, *ça va sans dire* in their own techno-ontology. SRs are predefined by their task-related and coherent-by-design functionalities. Accordingly, it seems that the appearance of the robotic otherness can only occur either in the design error, or in an unforeseen use of the technological object, which reveals an otherness devoid of instrumental functionality. Concluding, technological artifacts pose a double-trouble in the emergence of otherness, since they are both epistemologically predetermined, domesticated by the subject, and ontologically from the design process, in which sociality is confined in an instrumental intersubjectivity and anthropocentric[5] being-in-the-world. This double-faced perspective of otherness – experienced in the subjectivity of humans and in the designed objectivity of the SRs – we have tried to fix in the so-called Δ phenomenology of HSRI. ***When social robotics stops to be experienced in the subjective forms of socio-technical systems and, consequently, some of its task-related and coherent-by-design functionalities lose the meaning of a the simulated intersubjectivity, the otherness of the robotic technology acquires the human-decentred character of alienation, as we will show in the next chapter.***

Second block of considerations

Against this backdrop, we accordingly assume that

[5] This is also pointed out by both Coeckelbergh and Gunkel when they criticize Levinas' anthropocentrism in defining the "encounter", though they approach this problem only from the epistemic standpoint. The peculiar issue of robotics is that this anthropocentrism comes even before the interaction itself, in the design process.

- The more the task-related functionalities that ontologically designed SRs will adapt to the epistemological predetermination of the human subject, the less "otherness" can be experienced in HSRI.
- The more $\Delta(y,z)$ tends to zero[6], the less the robot is an "other" with whom I am committed to having feelings.
- It is precisely when the form of sociality is perfectly simulated by the coherent-by-design functionalities of the robots that we lose something, namely the experience of otherness[7].
- *We can therefore claim that an excess of task-related efficiency in robots' interactive functionality will end up reducing their effectiveness in producing a human-like relation, since the condition of possibility of "otherness" (Δ_{yz}) are eradicated.*

5. Alienation and recognition in the experience of HSRI

We have analysed HSRI from the side of subjectivity and objectivity. The relationship of intersubjectivity remains to be investigated in order to respond to the issue of whether HSRI contains possibility of mutual recognition. This account is particularly relevant to be understood in the human-CRs relationships because, if not, we may implicitly assume that is the same relational structure of HHI. The problem of the otherness of the other is therefore not only to be investigated from the point of view of the phenomenological perception of the subject, a solipsistic position, but from both sides of the relation. If we examine the problem of recognition from the centre of HRI, we are in the right perspective to engage with the problem of mutuality, namely the mutual recognition of other's alterity. Therefore, we are asking how the machine – in our case CR – engages in relationships with humans. With Mark Coeckelbergh and David Gunkel, in a Heideggerian fashion, we may "dispose" of the object or let otherness alter ourselves. On the other hand, however, it does not seem possible

[6] This means that there is no difference between the coherent-by-design sociality simulated by the robot and the meaning of sociality that x receives or attributes by acting with y.

[7] Accordingly, we could argue that Seibt's question (2017) may be then formulated this way: "what, if anything, is lost *because* the process is perfectly simulated?"

for the robot to recognize the otherness of the human being. As Arnaud Ramey (2005) would say “a chair is there for my sitting, but I am nothing from the perspective of a chair”. Therefore, in Kojeve’s words, but also in the later receptions of his work done by Judith Butler (2012), the robot lacks the desire for the other’s desire, namely the need to be recognized by another entity. In Kojeve, the other’s alterity emerges when this other formulates the recognition question: I am altered by the other when I am hit by the other’s demand for recognition. Although the human being can take an attitude that "lets otherness appear", a robot is neither able to receive this opening nor to reciprocate, since machines cannot ask for recognition. Therefore, this encounter will be only fictionally intersubjective, since the only relational actor is the human subject. A robot is a fictional “other” since it is a “you” without an “I”.

Nevertheless, CRs fictionally enact an intersubjective relationship, where the robot seems to be an “other” to which I relate, to the point that CRs seems to (at least partly) substitute human-human relationships (Turkle 2006). While in CRs this issue is clearly visible, these considerations can be extended to Social Robotics in general, since for the user they produce meaningful relations too. Moreover, most of the literature attests that humans relating with SRs actually threat them in a human-like way: they create emotional bonds, they get upset if interaction is not satisfactory; in short humans tend to anthropomorphize robots and treat them as if they were humans (Jung & Kopp, 2003; Konok et al., 2018; Krämer et al., 2011; Rosenthal-Von Der Pütten et al., 2014). This anthropomorphising passes through clear behavioural cues that users show when interacting with technological artefacts: for example they respond politely and behave differently with male or female voices (Reeves & Nass 1996) Then, what happens when an intersubjective relation is taking place from the standpoint of the human subject, but is only simulated from the other side? What will the human being receive in response to her demand for recognition? She will receive a confirmation of her epistemic pre-determination of intersubjective relations. While in HHI the other troubles my standpoint, producing an alteration of my point of view in the intersubjective relation, the robot cannot.

Thus, what am I experiencing in terms of intersubjectivity and recognition in the HSRI? We suggest that the ***robot may become the space of a displaced self-production of the human subject, which maintains a relationship with a robot that we can at this point define as a relational mirror***. As in the myth of Narcissus finding himself mirrored in the lake, the relational robot does not bring an otherness that alters the subjective standpoint of the human being. At most, it can re-enact from the "other" standpoint the meaning attributed by the human user to the relation. In Gunkel's words, the robot will actively confirm the validity of the subject's pre-determined conception of that particular relation: the meaning attribution about the relation made by the human subject. The robot otherness I experience is my self-subjectiveness displaced and alienated in the fictional otherness of the robot: the alterity I experience is myself in the other. This will be further explained in the next chapter. For now, using Ihde's graphical representation of interaction with a quasi-other, human-SR intersubjective relations may be described as:

Human → SocialRobot(~~myself~~)

"Myself" is crossed out to emphasize the fact that the displacement of oneself is concealed, that is the process of alienation. Continuing the reasoning of the previous chapters, ***a greater adherence on the part of the robot to human semantic attribution resulted in the greater task-related effectiveness of robot sociality; on the contrary, in supposed intersubjective HSRI, the more the robot's adheres to human's pre-categorized characteristics of otherness, the less alterity is experienced***. In conclusion, we claim that intersubjectivity with robots is a false flag even in presence of an "alterity relation": no intersubjectivity is possible since the robot cannot reciprocate the recognition process made by the human subject.

From the previous considerations, some conclusions can be deduced on the side of human relationship quality.

5.1 Alienation and recognition as manifestation of Δ phenomenology

Firstly, we can now argue that users can perceive the robot as an "other", and therefore develop a simulated intersubjective relationship, when their semantic attribution of robot sociality is overdetermined in respect to the actual task-related social functionality of the robot. In the language of the Δ phenomenology, ***the object of alienation in HSRI is the differential between "y" and "z", namely that hallucinatory and alienated part of the relationship completely produced by the human subject***.

human → SocialRobot (~~alienated-myself~~)
human → SocialRobot (Δy,z)

Here the differential (Δy,z) represents the remnant of significance that the human subject believes comes from the robot, while it is a displacement of herself. Under this respect, Δ can contain a large spectrum of psychological-relational consequences of HSRI. Although social robotics today might be presented as relational technology that allows forms of sociality and intersubjectivity, according to our differential (Δ) approach, the latter leads to alienation: what the human user finds in the robot's otherness is herself, but displaced as it were another.

5.2 Alienation, deception, and self-deception

A second set of conclusions refers to the philosophical problem known in the HSRI literature as "deception objection" (Stahl and Coeckelbergh 2016; Whitby 2008; Sharkey and Sharkey 2006). In our argumentation, the human → SocialRobot (Δy,z) overcomes the issue of robotic "deception". To date, robotic technology makes the "deception objection" arguably a detached discussion from current technology, as no robot is capable of deceiving a human about its artificial nature. Furthermore, the deception objection does not take into consideration the phenomena of self-deception, certainly much more likely to happen with current robotic technologies: users may voluntarily suspend their judgment on the more or less artificial nature of the robot during the interaction. Sexual robots are a

striking example of this mechanism (Cox-George & Bewley 2018). We believe that the discussion of the ethical issues of deception misses the target and, above all, diverts the discussion from the relational implications that current SRs may produce.

5.3 Alienation and subject omnipotence

A third set of considerations refers to psychological aspects of HSRI. We highlighted the alienating nature of HSRI, and accordingly we argue that such relationships have serious consequences on the psychological and relational health of individuals. It is not the aim of this paper to go into detail of the psychological implications of the subjective alienation process of HSRI. Nevertheless, we can identify one main psychological consequence of alienated human-robot relationships. If the subject fills both positions of the intersubjective relationship, this stands for a confirmation of subjective omnipotence. In fact, the robot covers a median position between objectivity and subjectivity: it is behaviourally able to display relationality, like a subject. On the other hand, the robot is "available" to the user in the same way as an object. In fact, robots do not manifest a relational standpoint: as Arnaud Ramey says it is not "interested" in the relationship. Therefore, a transfer into intersubjective relationships of the typical expectations of the object relationship may take place: the full availability of the object. This could lead users to transfer the same relational expectations of human-robot relationships in relationships with other human beings, especially in the case of vulnerable human groups such as elderly and children. In fact, these are the main users of current CRs. In conclusion, the dislocation of the subject in SRs may cause to some extent a disintegration of the self and its objectification, an obvious consequence of the processes of alienation.

Third block of considerations

- The robot double predetermination, epistemological and ontological, hinders the experience of a robot alterity. If the robot design perfectly adapts to the human epistemological pre-determination, namely $\Delta_{y,z} = 0$, no alterity may be experienced.
- There is no recognition possible in HSRI. What I can experience is the alienation of myself in a fictional other.
- The missing aspect in a perfectly simulated HSRI is the spontaneous emergence of the alterity and the mutual recognition of it. Despite social robotics may be perfectly interactive, humans are nothing from the perspective of the robot.
- When an alterity relation emerges, this sets up a hallucinatory and displaced self-production of myself in the fictional and alienated other.[8]

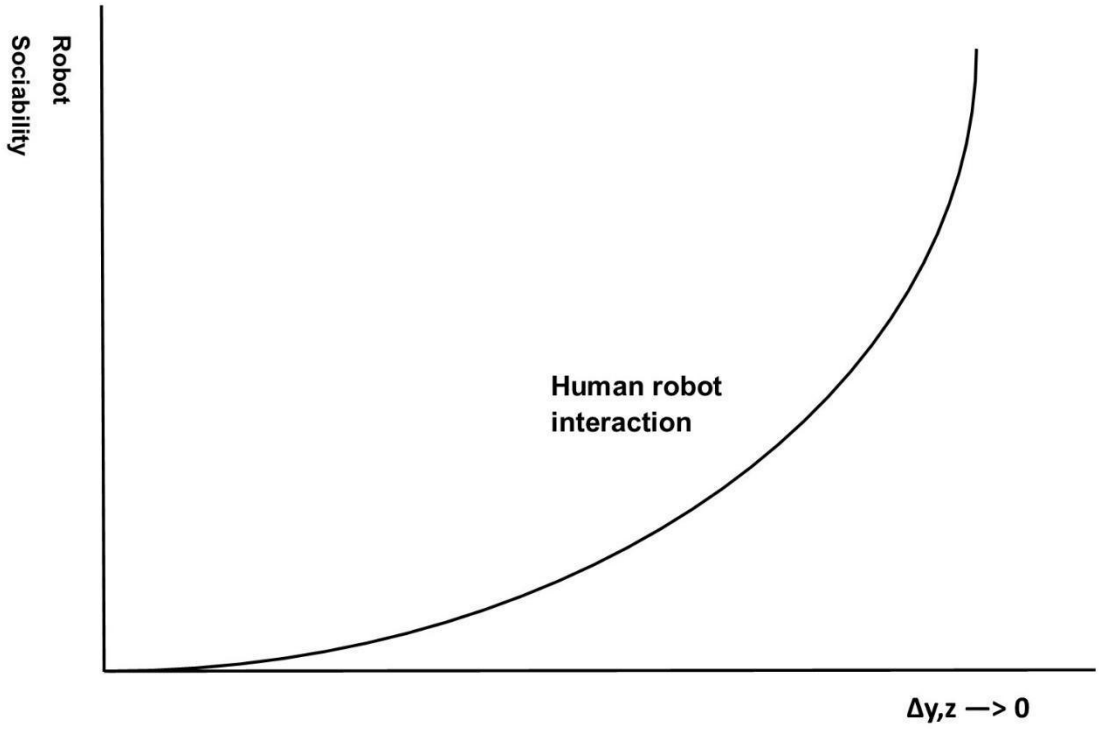
6. A robo-ethics for social robotics

The previous considerations allow us to make some further conclusions to intertwine the Δ phenomenology approach with robo-ethics (Veruggio & Operto 2008) and value-sensitive design (van Wynsberghe 2013). In our view, simulated intersubjective relationships between humans and robots can hardly lead to psychologically functional relationships. This can happen if and only if "y" and "z" are in the greatest possible overlap. The emergence of an otherness in the human-robot relationship, on the other hand, is exactly the space that opens up to dysfunctional and psychologically harmful relationships for the user. In order to ensure the functionality and ethics of HR relationships, it is necessary to reduce the gap between y and z, to ensure that the user semantic attribution of robot sociality and relationality is coherent with the actual robot socio-relational and task-related abilities.

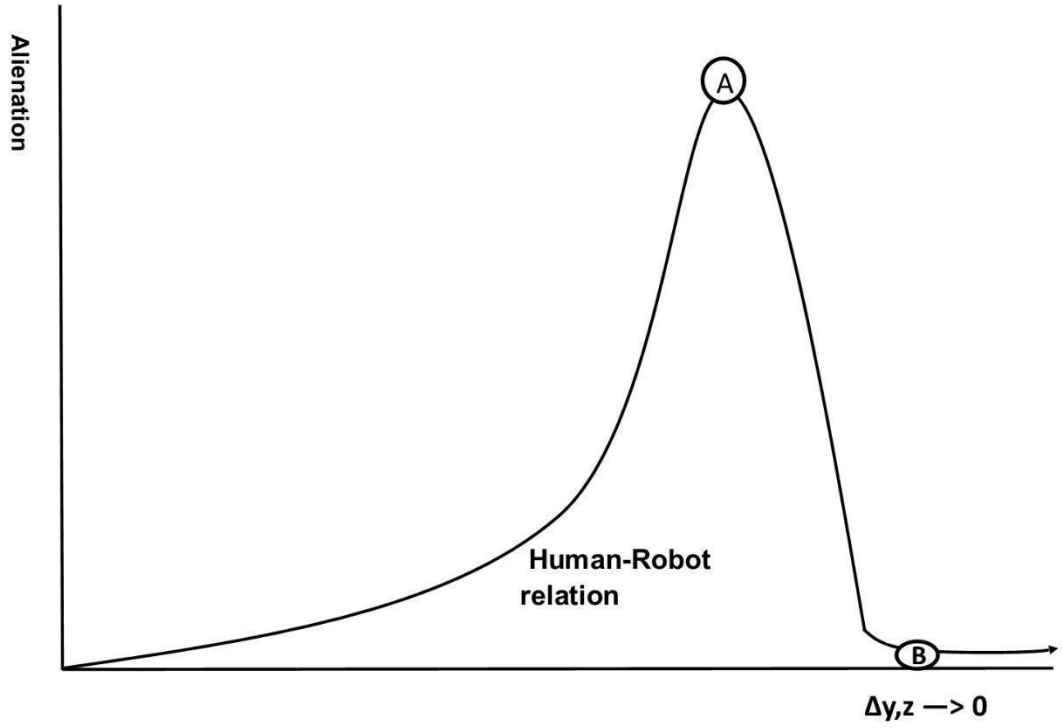
Therefore, a value-sensitive design recommendation develops around the search for the maximum possible sociality with the minimum possible incongruent semantic attribution, which ultimately

[8] This setting may bring users to hallucinate relations with robots, as Turkle highlighted in the elder relation with kismet (2006).

produce the alienating dynamic. To clarify the relationship between sociality and intersubjectivity, below we construct a graphical representation of their trends.



The sociality of the robot grows more and more as the differential (Δyz) decreases, as stated in the first chapter. In fact, at first the differential between the user semantic attribution and the robot's interactional abilities is too high to admit any form of social interaction. This simply results in the robot inability of being understood by the human interacting (Satake et al. 2009) When the differential decreases, a social bonding may be produced, and a social interaction starts.



As stated above, in the first phase, the robot social abilities are not enough to produce a social bonding, therefore there is little or no social interaction and there are no or little possibility that a user becomes alienated in the HR relation. When the differential starts to decrease, a social bonding is possible. In this moment, the alienating relation may be produced.

From the graphs we can see that:

- The intersubjective relation with a robot requires a certain degree of effectiveness in social interaction, otherwise there is no difference with an object-relation.
- The sociality of the robot grows more and more as the differential (Δyz) decreases.
- The intersubjectivity of the robot follows another trend: at first, the robot behaviour is too distant from user semantic attributions, therefore no social interaction takes place.
- Thanks to the progressive increase in interactionality, the degree of robotic intersubjectivity perceived by the user increases, until it reaches a peak. This represents the point where the user semantic attributions of meaning can be surreptitiously attributed to the robot behaviours.

Here the user is in the condition of maximum alienation in the robot. Obviously, this peak differs from user to user.

- When the differential between y and z approximates to zero, the degree of intersubjectivity of the robot collapses, as the semantic attributions operated by the subject returns to be congruent with the effective interactional capacities. In fact, the perfect adherence of the robot design to the predetermined epistemological structures of the human being leaves no room for the emergence of a fictional otherness.
- The point "A" displays the maximum point of alienation of the user, therefore the point of maximum dysfunctionality of the HSRI. A striking example can be the one reported by Turkle (2007). An elder of 76 years interacts with a My Real Baby doll: he gives to the doll his ex-wife's name, cuddles it, and takes it for long walks together.[9]
- "B" displays the breaking point where the relation becomes again psychologically functional.

From this graph we can deduce robo-ethical formulas that could regulate the value-sensitive design of SR:

Robot sociality function (s):

1. **f(s) = Δy,z —> 0**

HSRI alienation formula:

2. **human —> SocialRobot(myself)**

[9] The point A on the graph visibly recalls the peak of unlikability in the uncanny valley graphical representation (Mori 1970). We do not have space in this paper to extensively analyse the relations between the Δ phenomenology and the uncanny valley: a comprehensive discussion will be in next works. For now, we suggest that the uncanny effect and the alienation might be closely linked: the maximum point of alienation consists in a situation where the user experiences in the "other" a displaced self. If we follow the German word for uncanny, *unheimlich*, we discover that the uncanny feeling is due to something that is both home and non-home, familiar and non-familiar at the same time, as Freud already suggested (Freud 1919). Therefore, we tentatively claim that uncanny sensations might be the result of the process of alienation, in the sense explained above.

SocialRobot(myself) = Δy,z ≠ 0

human —> SocialRobot(Δy,z ≠ 0)

7. Conclusions

Moving from a semantic and ontological discussion on how to infer sociality in making experience of HSRI, the argumentation held in the paper has led to finally sustain three fundamental assumptions of a Δ phenomenology approach to robo-ethics. (1) The more the robot interactional functionalities are task-related, coherent-by-design, and congruent with a user's semantic attribution of sociality, the more the robot can be defined as social. (2) Forms of HSRI plainly designed to perform intersubjective relationships are actually particular forms of alienation, produced by the fictional intersubjectivity of the robot. This form of alienation increases when Δ increases: what is perceived as "otherness" in the simulated intersubjective HR relationship is the overdetermination of z on y, thus producing a structurally hallucinatory and dysfunctional relation. (3) Therefore, the only relationship with SRs that is simultaneously effective, psychologically functional and ethical is when the semantic attribution of the user is consistent with the real sociability of the robot. When we adopt such a constructive approach, the philosophical question of social robotics overcomes the ontological opposition between simulation and imitation of sociality and, more than this, becomes one of the phenomenologies constituting the socio-technical systems in which we live and by which we mediate our needs, desires, values. Our effort was to prove that, depending on the rationale we introduce into their sensitive design, social robots are still understandable as self-production of attributable meanings.

8. References

1. Asif, M., Sabeel, M., & Mujeeb-ur Rahman, K. Z. (2015, November). *Waiter robot–solution to restaurant automation. In Proceedings of the 1st student multi disciplinary research conference (MDSRC)*, At Wah, Pakistan (pp. 14-15)
2. Brandom, R. B. (2000). *Articulating reasons: an introduction to inferentialism*. Cambridge, MA: Harvard University Press.
3. Breazeal C., Dautenhahn K., Kanda T. (2016) Social Robotics. In: Siciliano B., Khatib O. (eds) *Springer Handbook of Robotics*. Springer Handbooks. Springer, Cham, pp 1935-1972.
4. Breazeal, C. (2002). *Designing Sociable Robots*. MIT Press.
5. Butler, J. (2012). *Subjects of desire: Hegelian reflections in twentieth-century in France*. Columbia University Press
6. Cassell, J., & Tartaro, A. (2007). Intersubjectivity in human–agent interaction. *Interaction studies*, 8(3), 391-410
7. Coeckelbergh, M. (2010). Health care, capabilities, and AI assistive technologies. *Ethical Theory and Moral Practice*, 13(2), 181–190.
8. Coeckelbergh, M. (2011). You, robot: on the linguistic construction of artificial others. *AI & society*, 26(1), 61-69
9. Coeckelbergh, M. (2016). Alterity Ex Machina. *The Changing Face of Alterity*, 181-96
10. Coeckelbergh, M. (2020). Monster Anthropologies and Technology: Machines, Cyborgs and other Techno-Anthropological Tools. in *Culture and Society*.
11. Cox-George, C., & Bewley, S. (2018). I, Sex Robot: the health implications of the sex robot industry. *BMJ Sexual & Reproductive Health*, 44(3), 161–164. https://doi.org/10.1136/bmjsrh-2017-200012
12. Freud, S. (1919). The Uncanny. *Imago*
13. Gunkel, D. J. (2007). Thinking otherwise: Ethics, technology and other subjects. *Ethics and Information Technology*, 9(3), 165-177
14. Gunkel D. J. (2016) Another Alterity. *The Changing Face of Alterity, 197-2018*

15. Gunkel, D. J. (2018). *Robot rights*. MIT Press

16. Dautenhahn, K. (2014). Human-Robot Interaction. In M. Soegaard & R. F. Dam (Eds.), *The Encyclopaedia of Human-Computer Interaction* (2nd ed.). Aarhus, Denmark

17. Grice, P. (1957). Meaning. *The Philosophical Review*, 64, 377–88.

18. Haikonen P.O.A. (2013) Consciousness and the Quest for Sentient Robots. In: Chella A., Pirrone R., Sorbello R., Jóhannsdóttir K. (eds) *Biologically Inspired Cognitive Architectures* 2012. Springer, Berlin, Heidelberg, pp. 19-27.

19. Hakli, R. (2014). Social robots and social interaction. In J. Seibt, R. Hakli, & M. Nørskov (Eds.), *Sociable robots and the future of social relations: Proceedings of Robo-Philosophy* 2014 (Vol. 273, pp. 105–115). IOS Press.

20. Hegel, G. W. F. (1807). *The phenomenology of spirit.*

21. Hegel, F., Gieselmann, S., Peters, A., Holthaus, P., & Wrede, B. (2011). Towards a typology of meaningful signals and cues in social robotics. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 72–78. https://doi.org/10.1109/ROMAN.2011.6005246

22. Honneth, A. (2004). Recognition and justice: Outline of a plural theory of justice. *Acta Sociologica*, 47(4), 351-364.

23. Honneth, A. (2007). *Reification: A Recognition-Theoretical View*. Oxford University Press.

24. Ihde, D. (1990). *Technology and the lifeworld: From garden to earth.* Indiana University Press.

25. Jaeggi R. (2014). *Alienation*. New York: Columbia University Press.

26. Kahn, P. H., Friedman, B., Perez-Granados, D. R., & Freier, N. G. (2004). Robotic pets in the lives of preschool children. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1449–1452).

27. Kojève, A. (1980). Introduction to the Reading of Hegel. *Cornell University Press*

28. Konok, V., Korcsok, B., Miklósi, Á., & Gácsi, M. (2018). Should we love robots? – The most liked qualities of companion dogs and how they can be implemented in social robots. In *Computers in Human Behavior*, 80(November), 132–142. https://doi.org/10.1016/j.chb.2017.11.002
29. Krämer, N. C., Eimler, S., Von Der Pütten, A., & Payr, S. (2011). Theory of companions: What can theoretical models contribute to applications and understanding of human-robot interaction? *Applied Artificial Intelligence*, 25(6), 474–502. https://doi.org/10.1080/08839514.2011.587153
30. Mori, M. (1970). "The uncanny valley", Energy, vol. 7, no. 4.
31. Mumm, J., & Mutlu, B. (2011). Human-robot proxemics: Physical and psychological distancing in human-robot interaction. HRI 2011 - *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, 331–338. https://doi.org/10.1145/1957656.1957786
32. Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Nonverbal leakage in robots: Communication of Intentions through Seemingly Unintentional Behavior. *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* - HRI '09, 2(1), 69. https://doi.org/10.1145/1514095.1514110
33. Ramey, C. H. (2005, July). For the sake of others": the personal ethics of human–android interaction. In *Proceedings of the CogSci 2005 workshop: toward social mechanisms of android science* (pp. 137-148).
34. Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.
35. Ricoeur, P. (1968). *Heidegger and the Quest for Truth*. Chicago: Quadrangle Books.
36. Rosenthal-Von Der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards

humans and robots using fMRI. *Computers in Human Behavior*, 33, 201–212. https://doi.org/10.1016/j.chb.2014.01.004

37. Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., & Hagita, N. (2009, March). How to approach humans? Strategies for social robots to initiate interaction. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 109-116)
38. Seibt J. (2017) Towards an Ontology of Simulated Social Interaction: Varieties of the "As If" for Robots and Humans. In: Hakli R., Seibt J. (eds) *Sociality and Normativity for Robots. Studies in the Philosophy of Sociality*. Springer, Cham.
39. Sharkey, A., & Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology*, 14(1), 27–40.
40. Sharkey, N., and Sharkey, A. 2010. The crying shame of robot nannies. An ethical appraisal. *Interaction Studies* 11 (2):161-190.
41. Siciliano, B., & Khatib, O. (Eds.). (2016). *Springer handbook of robotics*. Springer.
42. Sparrow, L., & Sparrow, R. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16(2), 141–161.
43. Stahl, B.C., and Coeckelbergh, M. (2016). Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems*, Volume 86, 152-161.
44. Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: the complexities of cybercompanionship. *Connection Science*, 18(4), 347–361. https://doi.org/10.1080/09540090600868912.
45. Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the 21st century. *Philosophy & Technology*(24), 251–268.
46. Jung, B., & Kopp, S. (2003). FlurMax: An interactive virtual agent for entertaining visitors in a hallway. *Lecture Notes in Artificial Intelligence*, 2792, 23–26. https://doi.org/10.1007/978-3-540-39396-2_5

47. van Wynsberghe, A. (2013) Designing Robots for Care: Care Centered Value-Sensitive Design. *Sci Eng Ethics* 19, 407–433.

48. Veruggio G., Operto F. (2008) Roboethics: Social and Ethical Implications of Robotics. In: Siciliano B., Khatib O. (eds) *Springer Handbook of Robotics*.

49. Whitby, B. (2008). Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. *Interacting with Computers*, Volume 20, Issue 3, 326-333.

50. Wood, A. W. (2004). *Karl Marx*. London: Routledge.