Companion Robots:
the Hallucinatory Danger of Human-Robot Interactions

Piercosma Bisconti Lucidi[1], Daniele Nardi[2]

[1]Sapienza University of Rome, Department of Philosophy, Via Carlo Fea, 2, Rome, Italy, piercosmabisconti@gmail.com

[2] Dipartimento di Ingegneria Informatica, Automatica e Gestionale "A. Ruberti", "Sapienza" Università di Roma, Via Ariosto 25, 00185 Rome, Italy, nardi@dis.uniroma1.it

Abstract

The advent of the so-called Companion Robots is raising many ethical concerns among scholars and in the public opinion. Focusing mainly on robots caring for the elderly, in this paper we analyze these concerns to distinguish which are directly ascribable to robotic, and which are instead pre-existent. One of these is the "deception objection", namely the ethical unacceptability of deceiving the user about the simulated nature of the robot's behaviors. We argue on the inconsistency of this charge, as today formulated. After that, we underline the risk, for human-robot interaction, to become a hallucinatory relation where the human would subjectify the robot in a dynamic of meaning-overload. Finally, we analyze the definition of "quasi-other" relating to the notion of "uncanny". The goal of this paper is to argue that the main concern about Companion Robots is the simulation of a human-like interaction in the absence of an autonomous robotic horizon of meaning. In addition, that absence could lead the human to build a hallucinatory reality based on the relation with the robot.

Five Ethical Concerns for Companion Robots

In recent years, advances in the field of robotics have led to a great expectation of what we call Companion Robots, namely machines that are designed and specifically programmed to produce a physic, linguistic and in some measure also emotional interaction with human beings. The main goal of these machines is often the care of children, elderly, people with disabilities. In addition to a general skepticism about the possibility of a positive reaction of public opinion to the introduction of such machines into everyday life (Coeckelbergh et al. 2016), other ethical con-

cerns have been raised. The questions posed by the human-robot interaction highlight the ethically correct modalities of this relation. They also raise ancient problems of philosophy in the field of robotics. In this first section we highlight the five main ethical concerns for Companion Robot (CR). In the second section we discuss the "deception objection", arguing that this ethical concern is based on a false opposition between human-human and human-robot interactions. Then, we argue that a risk for human-robot interaction is to become hallucinatory for humans. In the third section we analyze the definition of CRs as "quasi-others" in relation to the notion of "uncanny" in Freud (1919).

From the various scientific articles that discuss the future use of CRs, one can extrapolate five major criticisms or perplexities concerning scholars (Sharkey 2012). We will summarize and comment on these main topics. Attention will be given to the deception accusation, which will be reformulated from a different view of intersubjective relations. These five points specifically refer to robots for the care of the elderly, but are easily extendable to any CR. It is necessary, in commenting on these five most critical points, to distinguish between the undesirable consequences derived from CRs directly, and the unwanted consequences, where CRs are to be considered solely the incidental cause that triggers a deeper root cause. When discussing the effects of "Robotic Invasion", one should consider the difference between the changes that robotic causes *suo motu* and those that were already present within our time as tendencies, social or individual, that robotics helps to manifest. This distinction makes no less relevant the discussion on how to minimize the impact that robots might have on social, cultural, and political processes already in

place, a central issue for the success in the robot's inclusion in millions of people everyday life. But a greater theoretical depth is needed to understand the different values of human life changes that the machine will bring in the coming years. The five major concerns are:

1) The potential reduction of human contact: the likely progressive loss of interaction with other human beings of those who will be subject to robot-caring.

2) An increase in the sensation of objectification: robot care could increase the sensation of loss of active subjectivity.

3) A loss of privacy due to constant monitoring.

4) Loss of personal freedom.

5) Deception and infantilization: the creation of robots with the ability to simulate mental states, to interact verbally (and not) with humans, to understand and respond to their emotional states, raises the question of whether it is legitimate to deceive human subjects on the simulated nature of intersubjective robotic interaction. Moreover, an interaction only with robots could lead the subject to a state of infantilization.

According to the distinction made between the direct and indirect consequences on these objections to robotics, we believe that points 1,3 and 4 can be attributed to the first group of problems raised by robots, namely those that do not specifically address the nature of CRs behavior, but derive from an extension to the machine of socio-cultural problems already present in our societies.

1) The potential reduction of human contact is not a consequence of the CRs, but is caused by the general indifference that the young and adult sections of the population feel towards the elderly, who are already marginalized in closed environments such as nursing homes for the elderly, with a minimal interaction with the rest of the world, fundamentally limited to the staff of the facility. It is reported by Kanamori (2002), that interaction with the robot stimulates the sociality of elderly. The loneliness of seniors, abandoned by relatives, is a process of social and generational nature that has little to do with ethical problems of robotic.

3) Also the loss of privacy is a problem that most of the existing technologies already pose to the user: computers, smart phones, smart homes, any environment monitoring tool, pacemakers. This, of course, does not reduce the critical importance of the privacy issue, but we do not consider this as a specific problem of robotics, since CRs do not bring any significant changes to the existing privacy question, but they represent only a further field of application.

4) Likewise, the loss of personal liberty equally affects robotics and human caring. Both the issue of privacy and freedom are raised with repeated insistence not because they are actually a novelty in the condition of elderly, but because the fact that these problems are raised by robots creates in the public opinion a stronger reaction, resulting

solely from mistrust and a great suspicion in the robotics. This negative feeling of robotics has been detected in 60% of Europeans, who are deeply opposed to CRs (Coeckelbergh et al. 2016). Obviously, the problem of social perception of the CRs phenomenon is not a secondary issue, as positive acceptance by the population is necessary to continue research and innovation in this field.

Hallucinatory Meaning Overload

In this article we intend to concentrate especially on points 2 and 5, namely the objectification and deception of the user by the CR. We argue that:

1) These two points should be considered as ethical issues directly raised by the robotics, which will bring substantial changes in our way of considering intersubjective relations.

2) Deception accusation, as it is formulated, lacks the target because it sublimates human intersubjective relations, creating a false polarization between human-human and human-robot interactions, where the former is true and the latter is false.

3) The risk of objectification of the elderly should overturn in the risk of indiscriminate subjectification of the objective world, which could produce a hallucinatory reality.

4) The hallucinatory use of objects by humans is not new, but the verbal, empathic and linguistic responsiveness of CRs produces an essential change in the hallucinatory subjectivation process of the object.

5) The root of this problem arises from a fundamental and still unbridgeable difference between human-human and human-robot relations: the absence, in the robot, of a meaning horizon.

Deception accusation is the objection that more than anything else is raised against CRs (Stahl and Coeckelbergh 2016), (Whitby 2008), (Sharkey and Sharkey 2006). The charge is: even though the robot can simulate an emotional state, relate verbally, and thus cause an emotional attachment from the human subject, all these robot's behaviors do not come from a real mental state, from a genuine emotional affect, but only from an algorithm that orders its behavior. The fact that the positive aspects of a CR, namely the greater possibility of interaction of the elderly, are based on a deception would, according to Sparrow (2006) and other, make CRs non-ethical: "What most of us want out of life is to be loved and cared for, and to have friends and companions, not just to believe that we are loved and cared for, and to believe that we have friends and companions, when in fact these beliefs are false".

Sparrow's words underline a radical difference between the robotic relation, where "care" is simulated and is therefore false, and the human relation, where the relation is not simulated and therefore "true". In fact, the separation between false and true is much slimmer, since in any human-

human relation or social interaction there is performativity of the relational behavior. That is, in our relations many attitudes and behaviors are coded by social rituals, religious secularized tradition, recurring forms of behavior; elements that escape our will, but define the nature of much of our interactions. The fact that our behavior is extremely overstructured by a set of meanings, that we neither understand nor control but yet practice, is an acquired result of cultural anthropology, philosophy and psychoanalysis. Cultural differences, and hence the difference in the horizons of meaning shared in a community, determine a huge part of what we consider to be right and wrong in an interaction. In a society that gives significant importance to family, a subject who cares about an elder could be considered an example of deception? Probably this question would be opposed by considering that the person who cares the elder, because influenced by social norms, is not in itself conscious of this and so his action is "genuine". We should conclude that what produces a deceptive caring is the consciousness of the extrinsic causation of the emotional state that leads us to care-giving. But this would mean that what makes us ethical in our behaviors is the ignorance that these are influenced by a horizon of cultural significance. Among other things, if we keep the idea that deception is mainly derived from the consciousness that our behavior is overstructured, the robot could not be accused of deceiving, in fact it is not aware of its behavior.

Obviously, all those who advance the accusation of deception would not agree with these consequences. But it seems impossible to keep the deception objection for the robots without falling in contradiction. If we accept, as it is reasonable, that human behaviors are always to some extent influenced by external elements that escape our immediate understanding as: unconscious reasons, cultural norms, survival of religious elements in civil practices, we must also accept that human relations themselves are based on a form of deceit. Indeed, one can argue that deceit is due to the consciousness of heterogeneity of the causes of our behavior, but one would say that the only ethical behavior is the ignorance of one's own causes, as seen above.

Or it can be argued that robotic behavior is false, because it is programmed, giving the feeling of total absence of spontaneity, but this presupposes that human behavior is spontaneous. In fact, we cannot in any way demonstrate that human behavior differs in this from the robotic one, because the elements to consider for explaining human behavior are almost infinite: psychological, environmental, physical, etc. You cannot put anybody twice in the same situation, to see if he would behave differently. We claim that human behavior is spontaneous, in the sense of "free from external normative influences" without having any evidence. We can say that it is precisely the unknowability of this aspect which makes us perceive the behavior of

humans as free. But then we are talking about the complexity of interaction: if a robot, when deciding its action, could takes such a large number of contextual elements as a human, then the same feeling of spontaneity would be perceived.

Third, one can deny that human behavior is influenced by external elements and could oppose to a fully hetero-generated robotic behavior a completely endo-generated human one: the set of affections that lead us to the caring of another person are true if and only if they are determined exclusively by our emotion towards this person, and this emotion does not come from causes outside the relation itself. This thesis seems to be hard to support. First, to argue that human behavior is by no means hetero-generated goes against all psychoanalysis and mass psychology, cultural anthropology, and the philosophy of the last century. Secondly, it would be a fallacious argument of the kind *obscurum per obscurius*: what makes human caring "true" is the emotions, which are true because they are not further questionable in the causes.

The deception objection brought to CR is based on the opposition of human-human and human-machine relations as the former of "true" caring and the latter of "false" caring. The underlying assumption of this opposition is that there is a pure form of human relation, free from deceptive aspects, that provides a measurement meter of the falsity of the human-robot relation. But this pure form of relation does not exist, because each interaction is always loaded with a vast set of elements that over-signify the relation and affect the behavior of the subject by external causes. A trivial example: a woman is extremely attached to an elderly person. She cares for her while not being related to kinship, but the elderly lady in question is very much like the woman's dead mother, in appearance and behaviors. The behavior of the spontaneous care-giving is therefore strongly influenced by this psychological aspect. Is not this an example of deception like the robotic one? She does not really care for the elder woman, she is only remembering her mother.

Each human relation is to some extent influenced by aspects outside the intersubjective relation itself. The accusation of deception for CRs therefore relies on a false assumption: that relations between humans are to some extent "true" since care-giving is due to an unselfish attention to the other subject. But this attention is not unselfish at all, in fact it extremely objectifies the other subject of relation. The care-giving action, to a certain extent, fulfills desires that are completely external to the relation itself. In this sense, we must turn the deception charge against humans, instead of CRs: humans act for reasons entirely different from "true caring".

Here obviously we do not want to extend to human the deception objection, but we want to show that this charge

against robots is based on a misunderstanding and a sublimation of human purpose in behaving.

So, regarding the deception objection moved from Sparrow and others we can conclude that:

- It is based on an implicit assumption that sublimates intersubjective relations between humans, considering these relations as "pure" and lacking any degree of deceit.

- It establishes, without justification, the place to measure the ethicality of relational behavior on the subject's intent rather than on its effect.

- It presupposes a surreptitious and sublimated subjectivity emptied of all the unconscious, cultural and normative aspects that determine the behavior of a subject.

- It does not consider to any extent the self-deception movement operated by the subject (Coeckelbergh 2010)

From the above, we conclude that human relations are to a certain extent influenced by factors external to the two subjects of the relation, and to the relation itself. We will call the "overload of meaning" the process in which: a subject "A", in a relation, places on an object or another subject "B" a set of meanings, wholly independent of "B" and the relation between "A" and "B", and for "B" not significant. These meanings are meaningful only for the author of the overload. Often this overload is due to a cultural reason: to match the idea of sexuality with an intense sense of guilt is a typical consequence of the religiosity of the last century. This overload of the meaning of the object can also be seen in (Turkle 2006): during an interview with an elder who had received a CR, he said he considers the robot as his ex-wife and treat it as if it was her. The process of overloading meaning on an object can become, as in this case, a sort of hallucination.

For "hallucination" we mean here only the process where a subject adds meaning to another subject (or object) of the relation, that is external to it.

Trying to figure out how the relation can be overloaded we propose three ways:

- Psychological: the object of the relation becomes the signifier of psychological meanings.

- Cultural: Overload is based on the horizon of cultural sense of the subject as a result of the legal, moral, religious, historical, etc. normative constructions.

- Experiential: the subject carries his/her experiences into the relation, modifying judgment on objects and subjects based on past experiences.

We distinguished between "object" relations, those that include a human subject and an inanimate object, and "intersubjective" relations, which include two interacting subjects of some kind. The first difference between the two is that in the object relation there is no resistance to the overload of meaning, while in the subjective relations there is some. To return to Turkle, the elder who over-signified the small robot as it was the ex-wife, found no resistance from the robot to this projection; a human subject would certain-

ly have resisted to some extent to this. The second difference is that object relations offer to the subject a limited number of interactions, as an inanimate body does not take any initiative, and the quality of this interaction is repetitive. In short: we get easily bored of objects. Other human beings, on the contrary, offer numerically superior and qualitatively better interactions.

Robotic is advancing rapidly to bridge the existing distance between objective and intersubjective relations with regard to quality and number of interactions: the perfect Companion Robot would be able to provide a number and a quality of interactions equal to, or even bigger than, a human. However, what is not yet considered is the first gap, namely: to what extent a CR should accept to be meaning overloaded by a human? Our thesis is that the real threat CRs will bring to relations is not to "deceive" the subject, but to bring healthy subjects to gradually detach themselves from a balanced and "healthy" relationality in favor of a hallucinatory intersubjectivity.

The robot, in fact, appears to be the perfect hallucinatory object as it provides the maximum relational ability, as a human, and the minimal resistance to objectification. Therefore, we should consider the danger that, in a future where CRs really have a relational ability comparable to the human one, humans will start to prefer the friendship of robots instead of a human one, as Turkle (2006) argues. In fact, the robot not only interacts like a human, but does it every time the human wants and stops when he wants. In addition, the robot does not withstand to a meaning-overload as a human would.

From Quasi-Otherness to Quasi-Sameness

To define the role of the robot in the intersubjective relation, Coeckelbergh (2011) uses the notion of "quasi-other" He opposes the robot as "quasi-other" to robot as a simple machine, arguing that the difference between the two relies basically on a linguistic device. The quasi-subjectivity or the reduction of the robot to a simple machine is a linguistic device. Some considerations can be added to Coeckelbergh "quasi-other" definition. The robot is "quasi-other" because of its simultaneous being object and subject within the relation with the human. It puts in question the limits between objectivity and subjectivity. "Quasi-other" is at the same time something that exceeds its objectual nature, and a subject missing a crucial aspect of subjectivity. In each of the two cases the robot is located in a border area, where its definition always refers to an excess or a deficiency, an imbalance. It is therefore necessary to ask what the robot exceeds in objectivity, and what is lacking in being subject. While it is rather simple to define how the CR exceeds objectivity, that is entertaining an intersubjective relation with humans, less clearly one can identify what characterizes it as a missing subject. In fact, it is not enough for CR to simulate the relational behavior of hu-

mans, it remains a "quasi-other" and it does not reach a complete subjectivity. According to Coeckelbergh, the "quasi-other" nature of a robot is a linguistic construction that takes place within the generative processes of meaning of the normative structures of language. He points out two basic factors for the definition of the robot as a "quasi-other": the first is the robot's responsiveness to relational stimuli, which allows the robot to entertain a conversation, the second is the use, to address the robot, of the linguistic form "you": When the robot becomes a "you", it slips linguistically from objectivity to quasi-subjectivity.

According to the distinction above, we could say that verbal and relational responsiveness defines the content of the object's excess, while subjectivity is ensured by the linguistic normativity of the linguistically built "you" for the robot.

But Coeckelbergh does not address a second challenge that comes from the notion of "quasi-other": why does the linguistic construction of the robot as a subject not seem to be stable like the construction of any other subject? It seems that the robot, even if theoretically capable of performing all human behaviors in the same way as a human, cannot overcome its "quasi-other" nature and reach the state of a fully "otherness". We think an answer to this question can be found in the Freudian notion of "perturbing" or "uncanny". We are interested in using this term in the German meaning of "Un-heimlich", which translates to un-familiar. What is uncanny, for Freud, is basically something that is in an area of contradiction between the familiar and the stranger: something stranger that turns out to be all at once familiar, or vice versa. The uncanny effect originates in literature mainly from stories that have as central theme the sosia or the automata. From these two it is clear how this sensation originates, in particular from narratives that collapse the demarcation line between subjectivity and objectivity: the sosia is for Freud a copy of the self that recalls the possibility of death and reduction to objectivity; the machine produces an uncanny effect in a situation where one is unable to clearly identify whether the other is a human or an automaton.

The two main lines of uncanny can thus be identified in the pair of foreign-familiar and subject-object oppositions.

The "uncanny valley" theory argues (Pollick 2010) that the perturbing sensation originates in an area where the robot has a very strong resemblance to human, being but not yet complete. Freud would also subscribe to this assertion, as the disturbing machine in Freud is exactly what goes beyond human behaviors to be just a machine, but it is still too deficient to be called "human." The perturbing effect fades away if the robot does not look similar enough to a human being, or if this resemblance is total. For a physical and visual interaction, the uncanny feeling ceases when the robot makes its movements as fluid as humans. For what concerns interaction with CRs, the need is also to simulate

a human-like linguistic interaction. In fact, a CR would have a considerable number of interactions with the human subject, of various kinds and within a prolonged period. The challenge, then, is not to build an intelligence that can manage a single human-like conversation; the goal is to build a machine capable of entertaining any kind of conversation, building a memory of previous conversations to learn the communicative modalities of the human subject and bring the interaction to a deeper level. This must be done in the same way as humans do: diversifying interpersonal approaches based on multiple factors, as the knowledge of the other person and hence his "horizon of meaning." For example, if Alfredo talks to two of his friends and both say "I'm depressed," Alfredo diversifies the meaning of the word "depressed", based on previous interactions with his two friends. One of them might have a clinical history of depression, the other might be simply speaking about a moment of sadness. The robot must likewise become capable of this diversification of meaning and this will only be possible by "learning" the other's horizon of meaning that, as we have seen before, is the result of the interaction of subjective (psychological and experiential) and external (cultural contexts) elements. The greatest difficulty will be to understand in what way, for a particular individual, the set of culturally normative aspects meet with the personal history of the individual, and create a specific sense horizon. From this point of view, a CR should not only have the ability to penetrate into the behavior of human subjects, but also to put this behavior into a meaningful cultural context. But this is not the only ability the robot has to gain, to actually match a human from the point of relational effectiveness. In fact, every human is not only able to fit within the horizon of meanings of other people, but he/she carries his own.

For these reasons, we believe that the uncanny sensation will be produced by CRs by two fundamental aspects. First, the absence of an autonomous horizon of meaning. Second, the universal and abstract cause of the robot's actions. During any dialogical relation with another human, it is assumed that in front of the phrase "I fell in love", the other person is able to give not only a syntactic meaning to the sentence. We assume that it produces a mental state in him that attributes to "in love" a set of feelings, emotions and reasonings. These differ for a certain measure from those of the speaking subject, generating a semantic gap from the same word. This difference in the attribution of meaning originates from the set of cultural conditioning, psychological aspects, and the experiences of the other subject, and gives us the feeling that we are talking with another particular subject, carrying his/her own and unrepeatable meanings. This also means that when the other subject speaks, he is proposing a re-elaboration of the elements of the discourse coming from his semantic field, constantly making *in fieri* the meaning of the speech. The

significance of a signifier is consistently "unstable" during dialogue by continuous re-meaning that takes place within the semantic field of the two speaking subjects. In contrast, robots do not provide this feeling of interacting with another subject.

When Orelia, a girl who is entrusted with the robot AIBO (Turkle 2006), is asked if the robot can love her and if she can love it, the little girl replies that the robot cannot love her, at most it can be programmed to show love, but surely she cannot love something that does not reciprocate.

Here we can find both the differences that characterize the human-human relation from the human-robot relation: for the little girl the robot can perform loving behavior but cannot love. Loving behavior does not have any autonomous meaning for the robot, it is a significant behavior only for the human, and hence "They will not love you if you love them." The absence of a robot's autonomous horizon of meaning, that makes its behavior significant, is the first element of Orelia mistrusts. The second is the fact that programming is needed for the robot to perform a certain behavior. The fact that robot acts in a-priori way, not regarding the particularity of the interaction, makes the machine a universal abstract, opposed to human being, whose behavior is always contextualized. Also, if the robot acts stochastically, the resulting behavior is not perceived by human as spontaneous and interested about the interaction. We mean: if the robot says "I love you", I know that he is not loving me, but he is programmed to say "I love you". These words mean nothing for it. The CR, in the intersubjective relation, becomes for the human a semantic mirror: he does not produce an autonomous horizon to signify words and actions, outside the one produced by the human subject. Even though assuming a CR capable of diversifying the sense of words and experiences, based on the horizon of meanings of the human subject it interacts with, it could not nevertheless generate its own. That would be the only way to give humans the feeling of actually dealing with a subject. It could be argued that the robot, if able to learn the horizon of meaning of a human, would have a way of signifying the words and experiences "mimetic" with respect to that of that human subject. We believe that this is even more risky: if the robot assumes this mimetic pose, it definitely becomes a mirror of the subject. As we said before, this would lead to an unlimited extension of the hallucinatory overload on the robot. To return to the elder projecting the ex-wife on robot, in this case a CR would not only passively assume this role, but would actively perform behaviors associated with the hallucinatory overload of the human subject. That would lead users to an even more serious alienation in relations and would create a world tailored to their hallucinations. The situation would not improve even if CRs had the opportunity to draw a common network of intersubjective experiences and build a kind of "robotic shared subjectivity": this would be a universal subjectivity detached from the experience of a specific robot and thus seen as procedural and universalistic, as for Orelia, who argues that the robot is programmed to love and cannot truly be in love.

Going back to the "uncanny" theme, we believe that this sensation is basically generated by the two elements discussed above: the lack of a robot's own horizon of meaning and the feeling that its behavior is a disembodied universal. First, the uncanny sensation is generated by having to deal with an externalized self-mirror: this resumes precisely the Freudian definition of Un-heimlich: the contemporaneity of familiarity, human's own horizon of meaning, and foreignness, the fact that it is mimed by the robot. Secondly, it is uncanny the awareness that the robot acts in a disembodied way: this is because of the contradiction between an intersubjective relation that is perceived as a relation between two individualities, and the abstract and disembodied nature of the robotic act, where behaviors are perceived as determined a-priori. The robot is, in this regard, more an "almost-the-same", than a "quasi-other" compared to the subject: it is absent in him the ability to generate a universe of meanings for words and experiences, so he takes a human one. To sum up, the definition of CR as uncanny comes from "quasi-otherness" in two ways: first, it is a quasi-subject, namely it lies on the boundary between subjectivity and objectivity. Secondly, it is simultaneously a quasi-me/another-from-me: a subject-object reflecting the human horizon of meanings. This makes robot interactions uncanny in two ways: first, the CR, like Freud's automata, questions the difference between alive and dead, subjective and objective, particular and universal. Secondly, mirroring the human it produces the contemporaneity of a familiarity and foreignness.

Conclusions

The risk of creating a hallucinatory reality for humans in human-robot interactions is something which deserves an in-depth investigation. This does not mean that CR should be regarded as a threat to humans and society, but it is necessary to build human-robot interactions in such a way to ensure keeping the human subject psychologically healthy. We have illustrated that the difference between humans and robots relies in the human ability to make the semantic gap between two horizons of meaning fruitful. Consequently, to avoid a hallucinatory result, the challenge is to simulate this mechanism in robots. We are currently attempting a new theoretical paradigm that uses Lacanian theory of *Das Ding* (Lacan 1959) to design a healthier management of human-robot interaction.

References

Coeckelbergh, M. 2010. Moral appearances: Emotions, robots, and human morality. Ethics and Information Technology 12: 235-241.

Coeckelbergh, M. 2011. You, robot: on the linguistic construction of artificial others. AI & Society 26 (1), 61(9).

Coeckelbergh, M. 2010. Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations. Studies in Ethics, Law, and Technology Vol. 4: Iss. 3, Article 2.

Coeckelbergh, M., Pop, C., Simut, R. et al. 2016. A Survey of Expectations About the Role of Robots in Robot-Assisted Therapy for Children with ASD: Ethical Acceptability, Trust, Sociability, Appearance, and Attachment. Science and Engineering Ethics 22: 47.

Di Paolo, E. 2003. Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop. Dynamical Systems Approach to Embodiment and Sociality, Advanced Knowledge International.

Dreyfus, H. 2007. Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian. *In Skillful Coping: Essays on the phenomenology of everyday perception and action.* Oxford University Press.

Freud, S. 1964. The Uncanny. In J. Strachey (Ed. and Trans.), The standard edition of the complete psychological works of Sigmund Freud (Vol. 17, pp. 218-253). London: Hogarth Press.

Gunkel, D.J. 2015. The Rights of Machines: Caring for Robotic Care-Givers. In *van Rysewyk S., Pontier M. (eds) Machine Medical Ethics. Intelligent Systems, Control and Automation: Science and Engineering*, vol 74. Springer.

Himma, K.E. 2009. Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to Be a Moral Agent? Ethics and Information Technology, Vol. 11, No. 1.

Kanamori, M., Suzuki, M., and Tanaka, M. 2002. Maintenance and improvement of quality of life among elderly patients using a pet-type robot. Japanese Journal of Geriatrics 39(2): 214–218.

Lacan, J. 1959. The ethics of psycoanalysis, The seminar of Jacques Lacan, Book VII. Routledge

Pollick, F.E. 2010. In Search of the Uncanny Valley. In Daras P., Ibarra O.M. User Centric Media. UCMEDIA. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 40. Springer, Berlin, Heidelberg.

Sharkey, A., and Sharkey, N. 2012. Granny and the robots: ethical issues in robot care for the elderly. Ethics and Information Technology 14 (1): 27-40.

Sharkey, N., and Sharkey, A. 2006. Artificial intelligence and natural magic. Artificial Intelligence Review 25: 9-19.

Sharkey, N., and Sharkey, A. 2010. The crying shame of robot nannies. An ethical appraisal. Interaction Studies 11 (2):161-190.

Shibata, T., and Tanie, K. 2001. Physical and affective interaction between human and mental commit robot. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation,* vol. 3: 2572-2577.

Sparrow, R., and Sparrow, L. 2006. In the hands of machines? The future of aged care. Minds and Machines 16 (2): 141-161.

Stahl, B.C., and Coeckelbergh, M. 2016. Ethics of healthcare robotics: Towards responsible research and innovation. Robotics and Autonomous Systems, Volume 86, 152-161.

Turkle, S., Taggart, W., Kidd, C. D., and Dasté, O. 2006. Relational artifacts with children and elders: The complexities of cybercompanionship. Connection Science 18(4): 347-361.

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.