International Journal of Social Robotics (2023) 15:211–220
https://doi.org/10.1007/s12369-022-00958-y

Can Robots have Personal Identity?

Marcos Alonso[1,2]

Accepted: 12 December 2022 / Published online: 12 January 2023

Abstract
This article attempts to answer the question of whether robots can have personal identity. In recent years, and due to the numerous and rapid technological advances, the discussion around the ethical implications of Artificial Intelligence, Artificial Agents or simply Robots, has gained great importance. However, this reflection has almost always focused on problems such as the moral status of these robots, their rights, their capabilities or the qualities that these robots should have to support such status or rights. In this paper I want to address a question that has been much less analyzed but which I consider crucial to this discussion on robot ethics: the possibility, or not, that robots have or will one day have personal identity. The importance of this question has to do with the role we normally assign to personal identity as central to morality. After posing the problem and exposing this relationship between identity and morality, I will engage in a discussion with the recent literature on personal identity by showing in what sense one could speak of personal identity in beings such as robots. This is followed by a discussion of some key texts in robot ethics that have touched on this problem, finally addressing some implications and possible objections. I finally give the tentative answer that robots could potentially have personal identity, given other cases and what we empirically know about robots and their foreseeable future.

Keywords Robot ethics · Personal identity · Moral status · Relational view

1 Introduction

For many scholars, as well as for figures in the technological world such as Bill Gates, the most important technological revolution of recent years is the revolution in the robotics industry.[1] The idea that usually underpins these considerations is that "robots in society will be as ubiquitous as computers are today" [27, 3]. While such predictions are always debatable, it is clear that the presence of robots in the human world has only grown in recent decades. For more than a century industry has introduced countless machines that have exponentially boosted its productivity, many of these machines being robots. But for some years now, robots have been gaining their place in fields other than factories, manufacturing facilities and other typical industrial environments. More and more robots are being introduced in hospitals, nursing homes and various clinics [1, 45]. It is also increasingly common to see robots in education [24], in the military [40], in humanitarian aid or in the entertainment sphere [35, 57]. Even in intimate places like our homes, cleaning robots like Roomba, or toy robots like AIBO, are becoming part of the landscape. And, as has been widely advocated, "the prevalence of such robots is expected to increase strongly over the next few decades" [58].

This growing coexistence of humans with robots has led several authors to question the ethical implications of this relationship. Are robots mere objects or tools without any moral significance? While the latter position is adopted by many, the appearance and behavior of robots, often similar to human appearance and behavior, makes these questions not so easily resolved. We might think that this issue, if it has relevance at all, will only truly appear in a future where robots look so much like humans that they are almost indistinguishable. However, here I will argue, supported by several of the most relevant authors on this subject, that in principle it

[1] "The emergence of the robotics industry is developing in much the same way that the computer business did 30 years ago" [18].

✉ Marcos Alonso
marcs.alonso@gmail.com

[1] Departamento de Salud Pública y Materno-Infantil, Facultad de Medicina, Universidad Complutense de Madrid, Madrid 28040, Spain

[2] Universidad Adolfo Ibanez Campus Viña Del Mar, Viña Del Mar, Valparaíso, Chile

would not be necessary to wait so long for robots' existence to have significant moral implications.

This would have to do mainly with the relationship that robots will establish -that they are already arguably beginning to establish- with humans. If robots start to become our 'artificial companions' (Floridi 63) [12], if we can come to speak of 'social robots' [3], then it is very plausible that the moral burden of our relationship with robots can no longer be ignored. As M. Coeckelbergh has argued in numerous papers, if robots come to be able to interact with us in a human-like way, then the question of robot ethics will simply become unavoidable.

This paper builds on this approach, assuming that metaphysical or ontological discussions about what is a human being and what is a robot, while relevant, do not concern the core of ethics. Ethics is concerned with the relations between beings, not (directly) with their ontological context (see [42], 131–133). What a being is, what it is made of, and the capacities to which this constitution gives rise, are of fundamental importance for the relations that it can actually carry out. But it is these relationships that, in the final analysis, decide moral significance or insignificance.

However, the discussion around robotic ethics has almost always focused on problems such as the moral status or rights of these robots. The literature on these issues is extensive, with some works being particularly relevant, as [2, 4, 5, 32, 56, 59]. In this article I want to address a much less analyzed topic, but one that I deem crucial. The central question of this article is: can robots have personal identity and therefore be connected in some manner with human morality? The importance of this question has to do with the role we normally understand personal identity has for morality. Only when we identify and are identified by others does the moral world make sense. Ethics is not just agency, but recognition [25]. As far as I know, this key question has hardly been addressed, and when it has been touched upon it has been only tangentially and in a secondary way.[2] But before entering fully into the question of whether robots can come to have personal identity, it is useful to clarify this point about the relationship between identity and morality.

2 Identity and Morality

Personal identity, the fact that each of us considers and understands ourselves as unique beings, with characteristics that differentiate us from other human beings, and whose existence is prolonged in time, is essential for understanding our moral world. All our human practices, particularly those related to the moral world, are implicitly or explicitly based on personal identity. Only if I am the same as I was yesterday does it make sense to demand justice from the drunk who assaulted me in the bar. Only if I recognize my electrician -or at least recognize that person as an electrician- can I let him do his job quietly without calling the police to denounce this individual who is tampering with the wiring in my house. As Taylor [53, 48] explains, our personal identity allows us to orient ourselves in the moral world in the same way that our physical senses and capacities allow us to orient ourselves in the physical world.

However, the connection between identity and morality has been questioned by some authors. The most famous and influential questioning is undoubtedly that of D. Parfit in *Reasons and Persons*, who concluded in this famous work that personal identity is not what matters [37, 245–350]. But the most exhaustive critique of this connection between identity and morality has probably been that of D. Shoemaker [48]. For this author, the role we normally give to personal identity in our moral practices collides with the equivocality and confusion surrounding this concept. Shoemaker presents what he calls "the problem of multiplicity": the fact that judgments concerning personal identity occur in circumstances and contexts that are too diverse from one another, with very different practical demands and intentions. It is not the same, for example, to ask about who committed a crime, as to ask whether a former friend of ours, after a long time without seeing us, is still who we remember her to be. In this way Shoemaker tries to show the difficulties derived from understanding personal identity in a unitary sense, something ultimately impossible and which, therefore, should lead us to stop attaching so much importance to this concept [48, 354–355]

Although these criticisms are reasonable and interesting, they are based precisely on the recognition that is indeed given to personal identity in the moral sphere. While it is plausible to think, as Parfit or Shoemaker argue, that we should be more accurate and cautious with the use of this concept; the truth is that, for a large number of scholars, and I would dare say for a majority of people, there is a very strong and intuitive connection between personal identity and moral considerations. As far as this paper is concerned, we will operate under the idea that personal identity is key to

[2] There has been a recent debate on the so-called "electronic personhood" in the European Parliament [14], Prodhan, [39]). But I consider that this approach to personhood differs completely from what is raised here. First of all, personhood is not the same as personal identity. But even leaving aside this point that could perhaps be a mere terminological nuance, the key point is that "electronic personhood" is intended in analogy to the legal personality of companies and corporations, and this analogy, as Floridi and Mariarosaria [17] and Hubbard [21] have explained, is not accurate, since robots would not be persons like a company, but like a human being. Corporations depend directly on one or several humans, who act for them, and who are the ones who can decide to sell such corporations or dissolve them without any moral implications. To speak of the personal identity of a robot in a morally relevant sense, it could not be at the disposal of any human being in the same way as corporation.

morality as we usually understand it; and that therefore the discussion on robot ethics must include the discussion on the personal identity of these beings.

3 Personal Identity Theories

To give a better foundation to the subsequent discussion on robots´ personal identity, I would like to examine first, in a brief and panoramic way, the main theories of personal identity. In this way we will then be able to better delimit which conceptions could be applied to the case of robots, why and in what sense.

A first conception of personal identity is that defended by psychological theories of personal identity. These theories defend that, in order for a person to maintain his or her personal identity, there must be a psychological connection, a psychological linkage or a psychological unity between the different moments of consciousness or mental states of that person. This is the position usually attributed to Locke [30] and more recently recovered and reformulated by Parfit [37]. This position is the one that is usually understood to be intuitively defended by the bulk of the population [34]. The greatest virtue of this position is its ability to explain our everyday experience of what constitutes being oneself, that is, to sustain an auto-biographical account based on a memory and continued way of behaving over time. The major problem with these theories has to do with the fallibility of memory, which is often seen as the key to preserving this psychological unity and continuity. Parfit, for example, tries to overcome this difficulty by speaking of what he calls "quasi-memory" [37, 222], which would be the memory originated by the right cause, while giving much importance to personality and personal character. However, this way out is not entirely satisfactory and Parfit is even forced to conclude that personal identity is not really what matters, as noted above (for Parfit what really matters is what he calls relation-R [16]).

Secondly, we have the animalist theories of personal identity. According to these theories, our personal identity coincides with our body. Proponents of this position such as Olson [36] or Degrazia [63] understand that psychological theories err by taking the concept of personhood in a substantial, ontological sense, when its meaning is rather functional or descriptive [36, 69]. The strength of these animalist or biological theories lies in adequately reflecting many of our practices involving identity, such as identification through fingerprints, or through the photo on our identity card. The shortcomings of these theories have to do with their reductionism, leaving out key elements such as a person's personality, memories, projects, from personal identity. As Schechtman [42, 86] explains, in the typically human world, others are often presented to us as something more than "human organisms".

Finally, it is worth discussing narrative theories of personal identity. As the name suggests, these theories emphasize the narrative character of our personal identity, insisting on the fact that our identity is primarily a construction made by us and our human circle. Narrative theories do not disregard the importance of our corporeality in the shaping of our personal identity; but, unlike animalist theories, they consider that our corporal dimension must always be considered in connection with that auto-biographical dimension emphasized by psychological theories. However, in contrast to the individualism and solipsism of psychological theories, narrative theories emphasize the importance of our body, our institutions and material practices, and more generally of others, in shaping our identity. For narrative theories of personal identity, this co-constructed and relational character of our personal identity, the fact that our personal identity only emerges and can only be maintained in close and continuous contact with other human beings, is absolutely fundamental. As authors such as Lindemann [29] or Schechtman [42] explain, it is through others, through our interactions, our practices, our works and even our institutions, that this narrative identity is sustained and shaped in reality. As we will see below, this will be a key aspect for the discussion on the possibility of granting personal identity to robots.

4 Can Robots have Personal Identity?

Thus, after the previous sections we can now address our initial question of whether robots have or could ever possess a personal identity. As explained at the beginning of this paper, there is no doubt that this question is connected to issues such as what kind of entity the robot is and what capabilities it has or could have. These questions would seem to be of particular importance if we use the typology of personal identity theories that was just presented. Therefore, it would seem that animalist theories would reject on principle the possibility of ascribing a personal identity to a non-biological being such as a robot. For these theories, surely the question would only make sense if technology advanced so far as to create a synthetic being materially so similar to the human organism as to be almost indistinguishable. Psychological theories could surely not be so taxing; but ultimately their answer to our question would depend, equally, on a material problem of technological development: if computation were to advance to the point of creating an artificial mind so similar to the human mind that we could not properly distinguish it from an organic one, we could begin to speak of personal identity in robots.

As I began by stating in the article, I rather think that these ontological considerations are in the background, and that the ethical issue does not depend directly on them. As narrative theories of personal identity show, personal identity—and we could say the same of morality itself—takes place primarily on the relational plane.[3] If this is so, there would be an avenue for considering robots to have personal identity, even if their minds and bodies were not identical to humans. However, as narrative identity theorists themselves always warn, this position does not imply a "free bar" for personal identity and moral consideration. Both auto-biographical capacity and embodiment itself have a decisive weight in delimiting the relational possibilities of a given being. This and other adjacent issues are what I would like to explore in the remainder of this article. As initially stated, the importance of this personal identity conundrum for robot ethics is much higher that what the literature has shown up to this point, given that morality and personal identity are so strongly intertwined.

4.1 Potential and Limitations of the Relational Argument

One of the most prolific authors on robotic ethics has been M. Coeckelbergh, who has rightly also based much of his argumentation on the relational element that is generated between robots and humans. This author has criticized approaches focused on "robot rights" (the idea that robots could have moral or legal entitlements such as the right to body integrity or equality), arguing that the rhetoric of rights is too strong a form of moral consideration, and is not the most appropriate for the specific case of robots [7, 210]. Coeckelbergh argues for a "social-relational justification of moral consideration" [7, 210], a "roboethics (…) consciously anthropocentric" [6, 219], which leaves aside the ontological questions that normally occupy researchers in this area. The main motivations for Coeckelbergh to abandon this line of argument are that the arguments around the ontological basis of morality entail problems of high thresholds and delimitation of relevant features that are irresolvable; the unavoidable existence of the argument from marginal cases; as well as problems of determination and moral epistemology [8, 212]. As this author explains, this relational approach is much closer to how, in fact, robot developers think: "they care less about consciousness, more about (inter)action and what this does to us" [6, 219]. He therefore concludes that "instead of indulging in fantasies about moral robots with robot rights, we must be attentive to, and imagine, possibilities of living with personal robots that contribute to, and indeed co-constitute, good human lives in practice" [6, 221].

Coeckelbergh's approach, although very much in line with the proposal made here, surprisingly leaves aside the question of personal identity. His approach based on the relational aspect of robots would seem to demand the treatment of personal identity precisely as the basis of this relationality. But on this and other related issues, Coeckelbergh shows no interest, leaving a certain void in his argumentation. This author insists that we should not think of morality as something adhered to a certain entity, but that "instead, moral consideration is granted within a dynamic relation between humans and the entity under consideration" [8, 219]. But this approach, although suggestive and correct to a large extent, seems incomplete because it does not delve into the necessary conditions for this dynamic relation between a human being and another entity to actually take place (see [22] for some of these requisites, such as the ability to concretely elicit responses and cooperate on a certain level, like the ones involved in doing something as trivial as bringing someone a glass of water). Coeckelbergh pretends that the replacement of features by "features-as-experienced-by-us" [7, 214] phenomenologically resolves the question. But the problem of how, if at all, such a human–robot relationship can come about remains highly relevant, in my view.

Authors such as Kahn and his collaborators have pointed out, in this critical line, that the human–robot relationship must be examined from a logic of varying degrees of authenticity. These authors draw on Buber's distinction between two fundamental types of relationships: "I-You" and "I-It" [23, 379–380]. From this perspective, the double aspect, passive and active at the same time, of authentic relationships is emphasized. The Self needs a true You to become Self, and vice versa. In a similar vein, Setman [46] has recently argued for robots to sufficiently emulate the vulnerability and unpredictability of human beings before we introduce them in the human sphere. While this call for vulnerability and authenticity in the relationship is relevant, I am not sure that this line of argumentation does much to clarify when the relationship between robots and humans can be considered truly moral and when it cannot. That experience of the "You" that Kahn and collaborators talk about seems like something that could in principle be experienced from interactions with robots that are not fully conscious or active in a fully human sense. Pattison has explained that it is already commonplace for us to relate in a personal way to artifacts (e.g., by naming our cars) [38]. Other examples could be mentioned, from smartphones to baseball bats. So the problem seems to persist.

[3] Of course, a distinction could also be made here, as several authors have done (e.g. [15]) between agency and moral patience. I cannot elaborate on this distinction, its presuppositions and implications. But even if one were to adopt this terminology and conclude that robots could never become moral agents and could only be moral patients,even then, the point of our argument—that personal identity is central to morality—would remain undiminished.

4.2 Embodiment as a Limit of Relational Morality

M. Schechtman has discussed this and many other related problems at length in her book *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. In this work Schechtman proposes a theory of personal identity, the "person life view," in which she attempts to articulate certain principles of narrative theories with key ideas from psychological and animalist theories. Particularly valuable, in my view, is this author's attempt to defend a relational view of personal identity without neglecting the importance of the psychological and bodily characteristics of the subjects. My argument is supportive of this proposal, seeking a complicated balance between these elements.

As different researchers have argued (e.g., [21, 444], Schmiljun,[44], 76), a sufficiently human-like robot, in its appearance and behavior, should be considered human for all intents and purposes. To prove this, Danaher presents a hypothetical case. In this case, someone close to us, let's say our partner, suddenly reveals herself to be a robot. What Danaher argues is that it would be very strange if we would stop considering her a moral being just because of that revelation [9, 2032]. Of course, it would be very strange and surprising, and perhaps that secret would lead us to cut off our love relationship,but it would make no sense for us to stop considering as a moral being someone who had related to us in a fully human way. Schechtman posits, in an analogous sense, that if a being had the appearance and behavior of a human, that being would have to be treated as a human: it would have to be given a place in the human world, that is, it would deserve a person-space. In this author's terminology, "a nonhuman who does possess the forensic capacities is also capable of engaging with others in person-specific ways and so of living a person life within the social infrastructure that defines such a life" [42, 132]. In fact, this author will argue, leaving beings such as these out of person-space would constitute a case of oppression comparable to slavery or racism. Even if this is not an unproblematic analogy, to defend that the robot, from its design and even from its etymology ("robota" is a Czech noun, first used by Karel Čapek, that means "servitude", "forced labor") implies this condition of slavery, seems as little defensible as the argument that would pretend to attribute that same condition of slavery to any race or collective because of its origin. A relevant difference on this regard is that humans go through a developmental stage (infancy) where they are subject to a (justified) restriction of their freedom. Whether this line of argument would be applicable for at least some robots is up to debate; but even a paternalistic, freedom-restricting attitude towards robots would in itself entail that we are already engaging in a moral relation with them.[4]

Returning to the problem of whether robots could really engage in human relationships and participate in human space, Schechtman specifies that "the form of our person-specific interactions is deeply connected to facts about our embodiment" [42], 132), and that therefore "there will undoubtedly be some limitations on how different from human embodiment the embodiment of a nonhuman person can be" [42, 132]. In my view, here we encounter a key issue, which advocates of the relational stance in robot ethics often do not discuss sufficiently. Thus, we see how, as a boomerang, ontology rears its head again. However, as explained above, this ontology and these particular characteristics are only of importance insofar as they make possible, or not, the human relations that constitute morality.

Thus, the key to the moral consideration of robots lies in the relationship they enter into with humans. But these relational capacities are constrained by their very constitution. Not every form of embodiment allows, according to Schechtman, to participate in the human world and interact in a human way. However, the delimitation as to which embodiments allow one to participate in the human world and which embodiments do not is, for Schechtman, an empirical question that cannot be resolved in a purely theoretical way [42, 132]. Authors such as Torrance have defended the "Organic View" according to which there can only be organic persons and that the very concept of "artificial person" is a contradiction in terms [54]. Less bluntly, Schechtman thinks that sentience in particular would be a fundamental capacity for participating in human interactions. She suspects, although she does not dare to be definitive on this point, that robots could not properly develop this feature of sentience, and that this lack would therefore prevent them from engaging in authentically human relationships. Nevertheless, Schechtman, who a few pages earlier had defended the possibility of considering human beings in a vegetative state as possessing a certain personal identity [42, 77], has to leave the door open to the possibility that these non-sentient artificial persons might just end up being "strange persons" [42, 136].

If Schechtman is right and we are essentially dealing with an empirical problem, it is very interesting to look at some studies and experiments that have been carried out on this issue of the human–robot relationship. The experiments of psychologist S. Turkle have been showing for decades how children develop personal relationships with a wide variety of robots. Also according to this author's work, the elderly develop a notorious attachment to their robot caregivers, to the point of using them as intimate confidants [55, 109–15]. Kahn and collaborators have precisely addressed the question "can people engage substantively in reciprocal relationships with humanoids?" [23, 373]. To answer this question, Kahn and collaborators analyzed the explanations that preschool children provided about their behavior with the AIBO robot (a dog-shaped robot), and how this experience compared with

[4] I want to thank an anonymous reviewer for this idea.

their behavior with a stuffed dog. Among other results, a particularly significant one was that the children's attempts at reciprocity were almost four times more frequent with AIBO than with the stuffed dog [23, 375]. Studies by Hinds and coworkers [19] have also shed light on how these robot-human interactions or cooperations occur. Their findings are that human-like robots are treated kindlier and respectfully than mechanical-like robots [60, 159]. Another similar study, this time with soldiers, has shown that some soldiers feel emotionally connected to the anti-bomb robots that have saved their lives, even becoming saddened when they are destroyed [20, 49]. It has even been reported that people develop a strong sense of gratitude regarding the Roomba cleaning robot [43, 213–14]. All of which is in line with other research in which considerable empathy for robots has been observed, to the point that subjects hesitate to "kill" or "torture" them [10, 52].[5]

4.3 The Problem of Constructivism

Thus, it seems that most empirical evidence points to the fact that we can establish personal relationships with robots. The importance and degree of this relationship is debatable, and further empirical studies will be needed; but there are clear indications that such robot-human relationships are not impossible. If, as everything suggests, robotic engineering continues to advance and the similarity with respect to human beings, both in appearance and behavior, also increases, it would be logical that the answer to our question would be more and more in the affirmative. Robots can interact humanly with human beings, can occupy human roles, and therefore can have, at least to some extent, a personal identity. It would be a primarily relational personal identity, but from the perspective of the narrative theories of personal identity discussed above, this is already sufficient to argue for their inclusion in the human sphere. As also explained in previous sections, this implies that these beings with personal identity should also have some moral consideration—although the degree of this consideration is open to debate. I believe that, as disputable as this idea might be, just pointing out to this real problem of robot´s personal identity and its relevance for robot ethics make this these reflections valuable.

However, mistrust about this approach may persist for many. Does this mean that any object to which we conventionally grant a personal identity automatically enters our moral sphere? Would this move not imply falling into an untenable constructivism? A critic of this position would object that it does not follow from our tendency to anthropomorphize many of the objects with which we live that they can have moral status or consideration. Lindemann [28, 35–36] and the aforementioned Schechtman [42, 117–119] discuss this same problem in relation to the possibility of granting personal identity to children with hydrocephalus or elderly people with severe dementia. If we relationally grant personal identity to these humans, despite their cognitive abilities appearing so impaired, could not we consider granting personal identity to pets? Lindemann and Schechtman's answer is along the lines mentioned above, that the different embodiment of our pets -for example, our dogs- makes our interactions and expectations of these beings radically different from those of children with hydrocephalus or elderly people with severe dementia. For this reason, these authors explain, pets cannot be said to have a personal identity.

But this is precisely the point that strengthens the case of the robot. Robots are not just any object, because their appearance is human-like. Their behavior is also human-like. Experiments with them grant us invaluable information about ourselves [62]. That is why the accusation of constructivism is unfounded, since there are indeed elements in robots that justify this treatment and the relationship established with them. If the objection is that we are anthropomorphizing objects, that we are projecting characteristics onto these beings, the answer would be that, in fact, we always "anthropomorphize" others when we relate to them. This is what the CASA (Computers Are Social Actors) paradigm showed in the early 2000s through different experiments [33]. Strictly speaking, when we are in front of another human, their mental states, their capacity to feel and even their organic interior, are only assumptions that we do not verify. We are always projecting these characteristics onto the other—albeit on the basis that their appearance and behavior give indications in the same direction. If the appearance and behavior of robots allow, or even demand, these kinds of assumptions, we would have to admit that we are in a situation identical to the one we usually find ourselves in vis-à-vis other human beings.

5 Robotic Embodiment and Personal Identity: Some Issues

From what is discussed above, there does not seem to be a sufficiently solid theoretical objection to deny, a priori, the possibility of robots having personal identity. If we are really

[5] In relation to all these experiments, it is very interesting the appreciation of Scheutz who explains that: "while people, when asked explicitly, might deny that they think of the robot as a person, an animal, or an otherwise alive agent, this response generated at the conscious level might be forgotten at the subconscious level at which robots can affect humans so deeply. Social robots are clearly able to push our "Darwinian buttons", those mechanisms that evolution produced in our social brains to cope with the dynamics and complexities of social groups, mechanisms that automatically trigger inferences about other agents' mental states, beliefs, desires, and intentions" [43, 215–216].

facing an empirical question, we have no choice but to experiment and see how technological progress forces us (or not) to change our answers to these questions. We can, however, set up some mental experiments that will allow us to gain more clarity on these issues, assuming, however, that the answers we obtain in this way will always be precarious and provisional.

Some authors have already presented hypothetical cases that would serve to make some checks on the moral consideration that we in fact give to robots. In Sparrow's Turing Triage Test, this author argues that "machines will be people when we can't let them die without facing the same moral dilemma that we would when thinking about letting a human being die" [50, 307], and predicts this will not happen. This is primarily because, in this author's view, "machines would never be capable of the sort of embodied expressiveness required to establish a moral dilemma about "killing" a machine" [51, 306]. Despite the interest of this approach, I would argue that this conclusion is not very convincing, and that the posing of the dilemma is somewhat confused. There are already many cases of people who are extremely attached to objects, to the point of preferring them to people. That many of these people are unwilling to prefer the death of a stranger to the destruction of that object has more to do with (1) the legal consequences of translating that preference into action, (2) the ability to recover or reconstruct that object, a reversibility that does not exist with respect to human life. The first point is something that could eventually change, as has happened with countless legislative changes throughout history. The second point has to do with an aspect that we will see below and that could be relevant to this discussion.

One of the problems with the Sparrow case is that it puts us in an all-or-nothing situation. For it could be that robots have moral consideration, but to a lesser degree, as many think is the case with animals. Thus, preferring for a robot to "die" instead of a human, even if this preference were invariant and found no exceptions, would not be evidence that robots lack status or moral consideration. If we return for a second to the case of human beings in a vegetative state, it is clear that preferring their death to that of conscious human does not imply that the former have no moral status at all.

On the other hand, the Sparrow case does not faithfully represent the manifold and heterogeneous field of morality. Alongside mental experiments such as this author's, I would find it interesting to present other cases such as the one Levy presents regarding sex robots [26, 228]. If a romantic partner were to meet us with a sex robot, would they be offended in the same way as if finding us with a sex worker? Would she interpret it as a simple form of masturbation? If, in addition to a sex robot, this robot had the ability to talk and interact on other human levels, would the reaction of the spouse or romantic partner change? I would say that the answers to these questions would give many clues about the ability to attribute personal identity and moral consideration to robots. If our romantic partner sees the robot as a mere object, the response would expectedly be close to indifference; while, if the romantic partner believes or feels the robot has some sort of personal identity, the moral response would expectedly be much more noteworthy.

A very interesting criterion, also of a largely empirical nature, is the ability to establish long-term relationships, as argued by MacDorman and Cowley [31]. For these authors, this ability to maintain long-term relationships is what, for example, clearly differentiates the relationship we are able to establish with a dog -a being that differentiates us from other people, and that can remember us long after seeing us for the last time- and a robot. However, it is clear that this point can be reached by robots eventually. Even so, this allusion to temporality does contain a problematic point that I would not want to leave unaddressed.

One important point is that, as narrative theories of personal identity point out, our personal identity is deeply intertwined with our experience of temporality; with our understanding of the passage of time over days, years and decades, and with how this passage of time is reflected in our bodies. Robotic embodiment could make it extremely difficult to grasp this experience of temporality so central to our understanding of reality and our personal identities. As Hubbard [21, 448] highlights, the predictably more stable, and predictably more interchangeable or replenishable corporeality of robots could result in a personal identity substantially different from our own. This, however, is not a given, as any physical embodiment will degrade over time. And, remarkably, human organic embodiment is now degrading at a slower pace as medicine advances, pointing to a horizon where our body will deteriorate much slower, if at all [13].[6]

Perhaps even more relevant is the problem of the artificial, computerized mind, susceptible to extremely rapid and radical change [21, 449], susceptible to duplication [21, 432], and with a memory capacity incomparable to human memory [21, 449]. As MacDorman and Cowley argue, "to build a robot that lacks the ability to develop its identity and beliefs -or at least simulated beliefs- in tandem with evolving social relationships is to develop a robot that is stuck in a moment in time" [31, 381]. More concretely, the computerized, perfect memory a robot would have might prove too dissimilar to human memory and its imperfect, malleable and re-elaborative nature. If a robot had a perfect recall of every conversation and events, their personal identity might not be able to develop in a human-like manner.[7]

But perhaps the strongest contrast is found in the experience of that temporal limit that is death. What personal

[6] I want to thank an anonymous reviewer for this reflection.

[7] I want to thank an anonymous reviewer for this suggestion.

identity would a being have that, like the robot, lacks a notion of mortality? (Schmiljun [44], 76). It is possible that the only way for robots to have a sufficiently human-like personal identity involves introducing into their systems these notions of temporality and mortality that, due to their different constitution, they could not spontaneously generate on their own. Likewise, it may be that this "introduction" of these ideas or beliefs can only be generated through an imitation of what actually happens in humans, creating robots that gradually develop through human interactions [43, 218] and progressively grasping the human way of understanding reality in narrative terms [11, 71]. Also, creating robots that can -or even have to- die,although this may prove more difficult than first thought.

6 Conclusions

For some readers, these last objections, as well as others that may have been left out, demonstrate the inability of robots to possess personal identity. In my view, what these and similar objections show is that the personal identity of robots would have its own defining characteristics. It would undoubtedly be a personal identity different from the human one. But I do not think it would be sufficiently different, or different in the relevant respects, to be considered not properly a personal identity. For, as Hubbard explains, "though daunting, these problems should not be overemphasized. We manage to address issues of human personhood even though issues about the nature of the human mind and about human self-consciousness and identity are far from solved" [21, 428]. The problem of personal identity is a complex and contested issue, and it should be no less so in the case of robotic personal identity. In a sense, we know (or think we know) robot´s minds and functioning better than human´s, a knowledge that might distance ourselves from robots and prevent us from assigning them human-like morality or personal identity.[8] But this can be questioned in two ways. First, our knowledge of human biology is advancing very rapidly. Secondly, developments in robotics are getting increasingly complex, even reaching problems of emergent complexities no human can really grasp, as happens with some black-box AIs [47]. This means that we might get to a point where our robotics´ comprehension (or lack of it) might be on par with human biology comprehension (or lack of it). In a general sense, as Coeckelbergh has argued, we cannot place a much greater demand on robots than on humans [8, 238]. If, as has been explained, we are willing to grant personal identity to many borderline cases such as children with hydrocephalus or the elderly with dementia, robots should be judged under an analogous standard. In any case, what this discussion on robot´s personal identity also shows is that personal identity, and narrative identity particularly, is a highly contested issue that must be continuously revised and reconsidered. Throughout this article, the plausibility of robots possessing, or coming to possess, personal identity has been defended. As explained, this question of personal identity in robots, although hardly addressed by the robot ethics literature, is crucial for this field, since morality and personal identity are often understood as two inextricably linked domains. Finally, some possible objections to this approach have also been studied, concluding that none seem capable of denying, a priori, the possibility of robots having personal identity.

As many authors argue, discussions like this not only have importance in relation to the restricted field of robot ethics, but also allow us to improve our understanding of human morality itself [8, 240]. Nevertheless, the problem of personal identity in robots, and, in a broader sense, the field of robot ethics, is and will be in the coming years a field of utmost relevance. The improvement and growth in number of these human-like robots will only increase the urgency of these debates, which are not only about the moral consideration of these new beings, but also about the effect they will have on humans. Anticipating future discussions and providing some clarity on these complex issues is therefore a crucial task in which we must all participate.

Acknowledgements I want to thank the reviewers for their suggestions and insightful comments.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The Funding was provided by Fondo Nacional de Desarrollo Científico y Tecnológico (Grant numnber 11200050).

References

1. Alvey R (2021) Robotics in healthcare. Online Journal of Nursing Informatics. Online journal of nursing informatics, vol 25 (2)
2. Boscarato C (2011) Who is responsible for a robot's actions? In: van der Berg B, Klaming L (eds) Technologies on the stand: Legal and ethical questions in neuroscience and robotics. Wolfpublisher, pp 383–402

[8] I want to thank an anonymous reviewer for this insightful comment.

3. Breazeal C (2003) Toward sociable robots. Robot Auton Syst 42:167–175
4. Bryson JJ, Diamantis ME, Grant TD (2017) Of, for, and by the people: the legal lacuna of synthetic persons. Artif Intell Law 25:273–291
5. Calverley DJ (2006) Android science and animal rights, does an analogy exist? Connect Sci 18(4):403–417
6. Coeckelbergh M (2009) Personal robots, appearance, and human good: a methodological reflection on roboethics. Int J Soc Robot 1(3):217–221
7. Coeckelbergh M & Faculty of Behavioural, Management Social Sciences (2010) Robot rights? towards a social-relational justification of moral consideration. Ethics Inf Technol 12(3):209–221
8. Coeckelbergh M (2010) Moral appearances: emotions, robots, and human morality. Ethics Inf Technol. https://doi.org/10.1007/s10676-010-9221-y
9. Danaher J (2020) Welcoming Robots into the moral circle: a defence of ethical behaviourism. Sci Eng Ethics 26(4):2023–2049
10. Darling K, Nandy P, Breazeal C (2015) Empathic concern and the effect of stories in human-robot interaction. In: 2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 770–775
11. Dautenhahn K (2003) Roles of robots in human society—implications from research in autism therapy. Robotica 21:443–452
12. Dautenhahn K et al (2005) What is a robot companion—friend, assistant, or Butler? Intelligent robots and systems. In: IEEE/RSJ international conference on in intelligent robots and systems
13. DeGrazia, D. (2005). Human Identity and Bioethics Cambridge: Cambridge core. https://doi.org/10.1017/CBO9780511614484
14. de Grey AD (2004) Escape velocity: why the prospect of extreme human life extension matters now. PLoS Biol 2(6):723–726
15. Delvaux M (2016) Motion for a European Parliament resolution with recommendations to the Commission on Civil Law Rules on Robotics, (2015/2103(INL))
16. Gunkel D, Bryson J (2014) Introduction to the special issue on machine morality: the machine as moral agent and patient. Philos Technol 27(1):5–8
17. Ehring D (2013) Why Parfit did not go far enough. Philos Stud 165(1):133–149
18. Floridi L (2008) Artificial intelligence's new frontier: Artificial companions and the fourth revolution. Metaphilosophy, 39(4–5), 651–655
19. Floridi L, Taddeo M (2018) romans would have denied robots legal personhood. Nature 557(7705):309
20. Gates B (2007) A robot in every home. Sci Am 296(1):58–65
21. Hinds PJ, Roberts TL, Jones H (2004) Whose job is it anyway? a study of human-robot interaction in a collaborative task. Hum Comput Interact 19:151–181
22. Hsu J (2009) Real soldiers love their robot brethren. LiveScience, May 21. http://www.livescience.com/technology/090521-terminator-war.html (accessed March, 14, 2021)
23. Hubbard FP (2011) Do androids dream? personhood and intelligent artifacts. Temp. L. Rev. 83:405–474
24. Ivaldi, S, Lyubova, N, Gerardeaux-Viret, D, Droniou, A, Anzalone, SM, Chetouani, M, Sigaud, O. (2012). Perception and human interaction for developmental learning of objects and affordances. In: 2012 12th IEEE-RAS international conference on humanoid robots (Humanoids 2012), pp 248–254
25. Kahn P, Ishiguro H, Friedman B, Kanda T (2006) What is a human? toward psychological benchmarks in the field of human-robot interaction. In: ROMAN 2006 special session: Psychological benchmarks of human-robot interaction.
26. Lepuschitz W, Merdan M, Koppensteiner G, Balogh R, Obdržálek D (2021) Robotics in education: methodologies and technologies. In: Advances in intelligent systems and computing, vol 1316. Cham, Switzerland
27. Lévinas E (1987) Time and the other and additional essays. Duquesne University Press, Pittsburgh, PA
28. Levy D (2012) The ethics of robot prostitutes. In: Lin P, Abney K, Bekey G (eds) Robot ethics: the ethical and social implications of robotics (Intelligent robotics and autonomous agents). MIT Press, Cambridge, Mass
29. Lin P, Abney K, Bekey G (2012) Robot ethics: the ethical and social implications of robotics (Intelligent robotics and autonomous agents). MIT Press, Cambridge, Mass.
30. Lindemann H (2002) What child is this? Hastings Cent Rep 32(6):29–38
31. Lindeman H (2009) Holding on to Edmund: the relational work of identity. In: Lindemann H, Verkerk M, Walker M (eds) Naturalized bioethics: toward responsible knowing and practice. Cambridge University Press, New York, pp 65–79
32. Locke J (1690/1975) An essay concerning human understanding. Clarendon Press, Oxford
33. Macdorman K, Cowley S (2006) Long-term relationships as a benchmark for robot personhood. In: ROMAN 2006—The 15th IEEE international symposium on robot and human interactive communication, pp 378–383
34. Miller LF (2015) Granting automata human rights: challenge to a basis of full-rights privilege. Hum Rights Rev 16:369–391. https://doi.org/10.1007/sl12142-015-0387-x
35. Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. J Soc Issues 56(1):81–103
36. Nichols S, Bruno M (2010) Intuitions about personal identity: an empirical study. Philos Psychol 23:293–312
37. Nielsen J, Lund HH (2008) Modular robotics as a tool for education and entertainment. Comput Hum Behav 24(2):234–248
38. Olson E (1997) The human animal: personal identity without psychology. Oxford University Press, Oxford
39. Parfit D (1984) Reasons and persons. Clarendon, Oxford
40. Pattison G (2007) Seeing things: deepening relations with visual artefacts. SCM Press, London
41. Prodhan G (2016) Europe's Robots to Become 'Electronic Persons' Under Draft Plan, REUTERS (June 21, 2016), http://www.reuters.com/article/us-europe-robotics-lawmaking-idUSKCN0Z72AY [https://perma.cc/UH3P-PW5B]
42. Scharre P (2019) Army of none: Autonomous weapons and the future of war. W. W. Norton & Company, New York
43. Schechtman M (1996) The constitution of selves. Cornell University Press, Ithaca
44. Schechtman M (2014) Staying alive: personal identity, practical concerns, and the unity of a life. Oxford University Press, Oxford
45. Scheutz M (2012) The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey G (eds) Robot ethics: the ethical and social implications of robotics (Intelligent robotics and autonomous agents). MIT Press, Cambridge, Mass
46. Schmiljun A (2018) Robot morality: Bertram F Malle's concept of moral competence. Ethics Prog 8(2):69–79
47. Sequeira J (2019) Robotics in healthcare: field examples and challenges. In: Advances in experimental medicine and biology, vol 1170. Cham, Switzerland
48. Setman SA (2021) A willingness to be vulnerable: norm psychology and human–robot relationships. Ethics Inf Technol 23:815–824. https://doi.org/10.1007/s10676-021-09617-8
49. Shin D (2021) The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int J Hum Comput Stud 146:102551
50. Shoemaker D (2007) Personal identity and practical concerns. Mind 116(462):317–357
51. Singer P (2009) Wired for war: the robotics revolution and conflict in the 21st century. Penguin Press, New York

52. Sparrow R (2004) The turing triage test. Ethics Inf Technol 6(4):203–213
53. Sparrow R (2012) Can machines be people? reflections on the turing triage test. In: Lin P, Abney K, Bekey G (eds) Robot ethics: the ethical and social implications of robotics (Intelligent robotics and autonomous agents). MIT Press, Cambridge, Mass
54. Suzuki Y, Galli L, Ikeda A, Itakura S, Kitazaki M (2015) Measuring empathy for human and robot hand pain using electroencephalography. Sci Rep 5(1):15924
55. Taylor C (1989) Sources of the self. Cambridge University Press, Cambridge
56. Torrance S (2008) Ethics and consciousness in artificial agents. AI & Soc 22(4):495–521
57. Turkle S (2011) Alone together: why we expect more from technology and less from each other. Basic Books, New York, NY
58. Turner J (2019) Robot rules: regulating artificial intelligence. Cham, Switzerland
59. van Wynsberghe A, Comes T (2020) Drones in humanitarian contexts, robot ethics, and the human–robot interaction. Ethics Inf Technol 22:43–53. https://doi.org/10.1007/s10676-019-09514-1
60. Veruggio G (2006) EURON roboethics roadmap (release 1.1). EURON Roboethics Atelier, Genua.
61. Wallach W, Allen C (2009) Moral machines: teaching robots right from wrong. Oxford University Press, Oxford
62. Walters M, Syrdal L, Dautenhahn D, Te Boekhorst S, Koay K (2008) Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. Auton Robot 24(2):159–178
63. Wykowska A, Chaminade T, Cheng G (2016) Embodied artificial agents for understanding human social cognition. Philos Trans Biol Sci 371(1693):20150375

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Marcos Alonso is Assistant Professor of Bioethics at the Medicine Faculty of Complutense University of Madrid (Spain). He holds a PhD from the Complutense University of Madrid and has previously worked in Spain and Ecuador. He specializes in applied ethics and philosophical anthropology. He is also a scholar of Spanish philosopher Ortega y Gasset.