Philosophy & Technology (2024) 37:51
https://doi.org/10.1007/s13347-024-00737-9

RESEARCH ARTICLE

The Moral Status of Social Robots: A Pragmatic Approach

Paul Showler[1]

Received: 3 February 2024 / Accepted: 24 March 2024 / Published online: 9 April 2024

Abstract
Debates about the moral status of social robots (SRs) currently face a second-order, or metatheoretical impasse. On the one hand, moral individualists argue that the moral status of SRs depends on their possession of morally relevant properties. On the other hand, moral relationalists deny that we ought to attribute moral status on the basis of the properties that SRs instantiate, opting instead for other modes of reflection and critique. This paper develops and defends a pragmatic approach which aims to reconcile these two positions. The core of this proposal is that moral individualism and moral relationalism are best understood as distinct deliberative strategies for attributing moral status to SRs, and that both are worth preserving insofar as they answer to different kinds of practical problems that we face as moral agents.

Keywords Moral status · Social robots · Pragmatism · Ethics

1 Introduction

Advances in AI and engineering may soon revolutionize the boundaries of our moral communities. The effects of integrating social robots (SRs) into our homes, workplaces, schools, hospitals and labs, transportation sector, military, and entertainment industries will undoubtedly be far-reaching.[1] The more sophisticated these machines become—as they increasingly display intelligence, emotions, autonomy, creativity, and sapience—the deeper and more complex the relationships we are likely to form

[1] Following Kate Darling, I take a social robot to be "a physically embodied, autonomous agent that communicates and interacts with humans on a social level (Darling 2016, 2). Whether this definition includes chatbots and other large language models is an open question (given that these technologies are technically embodied in hardware). The pragmatic approach to moral status developed in this paper could, in principle, apply to these cases as well. Although I shall focus primarily on SRs given their centrality in recent debates.

✉ Paul Showler
paul.showler@sdsmt.edu

[1] Department of Humanities, Arts, and Social Sciences, South Dakota School of Mines and Technology, 501 E. Saint Joseph Street, Rapid City, SD, USA

with them.[2] Some philosophers have no trouble envisioning near-future scenarios in which we have extended moral consideration to intelligent machines (Bostrom, 2014; Floridi, 2002; Gunkel, 2014; Tavani, 2018). On this view, as robots acquire human-like capacities we shall owe them certain forms of respect and find ourselves weighing their interests against our own. Future engineers may come to have the kinds of obligations toward their robotic inventions that parents have toward their children (Schwitzgebel & Garza, 2015, 108–9).[3] Other philosophers doubt that these scenarios are likely to occur—at least anytime soon (Andreotta, 2021; Mosakas, 2021; Müller, 2021). For these skeptics, social robots are artifacts. No matter how complex their behavior becomes, our obligations to them will never differ significantly from those we have to our toasters. Of course, owing to our own psychological tendencies to project mentality or agency onto social robots, we may risk mistakenly ascribing moral patiency to them. But these are outcomes to be avoided, perhaps by implementing design principles which limit the degree to which machines are intended to physically resemble humans (Bryson, 2009).

The question of whether we could have moral obligations to SRs is part of a broader debate concerning their *moral status* or their *moral personhood*.[4] To say that an entity has moral status means that its interests matter for their own sake, or that it is worthy of being treated with moral concern.[5] While some authors use the terms moral status and moral personhood interchangeably, more commonly, the former is used to denote that an entity has intrinsic moral worth, whereas the latter indicates a higher level of moral considerability (Gordon & Gunkel, 2022; Kamm, 2007).

Philosophical debates about the moral status of social robots have encountered two significant impasses that show no signs of abating. The first, and more widely acknowledged impasse occurs at a first-order level, and concerns questions about the grounds of moral status of SRs. Theorists are divided about which set of properties—for example, sentience, intelligence, sapience, or some other feature—justifies our having moral obligations toward SRs. Some writers advocate for a unicriterial account of moral status, maintaining that only a single property can confer moral status to SRs. Others defend multicriterial approaches, according to which a variety of properties are morally relevant.

[2] For a general discussion of the potential impact of robots within our lives, see Bostrom (2014), Darling (2016; 2021), Nørskov (2016). For a discussion of the economic impact of integrating robots into the workplace, see Ford (2015), Danaher (2017). Some writers have considered features of human–robot relations, especially sexual and romantic relations with robots (Danaher 2019; Frank and Nyholm 2017; Jecker 2021a; McArthur 2017), but also friendship (Jecker 2021b; Marti 2010) and care-giving (Sharkey and Sharkey 2010). Since the European Parliament's Committee on Legal Affairs issued a 2017 report proposing the creation of the category of "electronic personhood," there has been considerable discussion of the legal status of social robots. For an overview of this debate, see Parviainen and Coeckelbergh (2021).

[3] In addition to the parent–child relationship, other relationships that have been used to analogize the human–robot relation include the employee-employer relation, or the god-creature relation.

[4] In this paper, I take the term *moral status* to be synonymous with 'moral standing', 'moral considerability' and 'moral patienthood'.

[5] For related characterizations of the concept of moral status, see Warren (1997), DeGrazia (2008, 183), DiSilvestro (2010, 12), and Harman (2003, 174). Jaworska and Tannenbaum (2013) offer an overview of the literature on the grounds of moral status.

This paper focuses on a distinct second-order impasse. Here, disagreement concerns not which properties ground moral status, but rather the viability of attributing moral status based on an entity's properties in the first place. I shall use the terms *moral individualism* (MI) and *moral relationalism* (MR) as labels for the two sides of this debate. Moral individualists contend that whether an entity has moral status is a function of its intrinsic properties (McMahan, 2005; Rachels, 2005).[6] Moral relationalists, by contrast, deny that inquiry into the first-order questions is necessary or fruitful. When it comes to deliberating about the scope of one's obligations, these writers argue that moral status attributions should not result from identifying status-conferring properties, but from some other type of reflection—for example, through practices that expand our imagination and emotional sensibilities (Coeckelbergh, 2010, 2012, 2014, 2018; Gunkel, 2011, 2014, 2018; Jecker et al., 2022a).

It is typically assumed that MI and MR are mutually exclusive positions.[7] The central aim of this paper is to contest this assumption by arguing that it is possible to embrace both views, albeit in constrained forms.

Section 2 examines recent debates concerning the moral status of SRs and argues that both the individualist and relational strategies encounter serious limitations. On the one hand, moral individualism faces problems involving semantic indeterminacy, as well as problems stemming from skepticism about the normative relevance of status-conferring properties. On the other hand, I contend that moral relationalism fails to account for the fact that appeals to such properties do often factor into moral justification.

In the face of these challenges, I contend that individualists and relationists have good reason to adopt constrained versions of their positions. A constrained individualism holds that a social robot's possession of salient properties can factor into our moral reason-giving practices. But it denies that any property can provide agent-neutral moral reasons. A constrained relationalism allows that there are certain practical contexts in which appeals to status-conferring properties are appropriate, while continuing to insist on the indispensability of alternative modes of moral reflection and deliberation for other practical contexts.

Section 3 presents the core argument for a pragmatic account of moral status, which aims to reconcile constrained versions of MI and MR. My primary aims are (1) to justify the claim that MI and MR can be synthesized, and (2) to show how they can be reconciled in practice—especially in cases where the two approaches appear to conflict.

Rather than view MI and MR as competing theories about the nature of moral status, a pragmatic approach conceives of them as distinct deliberative strategies for attributing moral status to SRs. I argue that because both strategies answer to different practical problems that moral agents face, they are valuable for different purposes. Moral individualism—when suitably constrained—presents a set of

[6] James Rachels describes moral individualism as "a thesis about the justification of judgements concerning how individuals may be treated. The basic idea is that how an individual may be treated is to be determined, not by considering his group memberships, but by considering his own particular characteristics" (Rachels 2005, 173).

[7] Recent exceptions include Gordon (2021) and Gordon and Gunkel (2022). I discuss their view below in Section 3.

deliberative strategies that provides value clarity and that proves helpful in solving coordination problems. By contrast, moral relationalism presents a set of deliberative strategies that preserve the complexity of our values. These complexity-preserving strategies are crucial for individual moral development and for facilitating moral transformation within communities. Elucidating the practical value of these deliberative strategies not only lends support to the claim that MI and MR can be synthesized (by showing that both approaches answer to separate practical concerns), but it also provides a blueprint that shows how they might be integrated in practice—a task I take up in Section 3.4.

2 The Moral Status of Social Robots: Current Views

2.1 Moral Individualism: For and Against the Moral Status of Social Robots

Most philosophers concerned with the moral status of SRs subscribe to moral individualism—or as it is often called—a property-based view (Andreotta, 2021, 24; Gordon, 2021, 463; Mosakas, 2021, 430). According to this position, whether a social robot has moral status depends on whether it possesses morally relevant intrinsic properties, such as sentience, sapience, intelligence, or rationality, among others.

Moral individualism invites (at least) three sets of questions: *metaphysical questions* about which properties are status conferring, *empirical questions* about whether social robots can (or do) instantiate these properties, and *epistemological questions* about whether we can know that SRs instantiate them. Given that these questions currently lack definitive answers, moral individualism is consistent with arguments both for and against the claim that social robots may someday possess moral status.

Consider the view—call it *robot rights optimism*—that social robots either currently, or will likely someday soon possess moral status.[8] Eric Schwitzgebel and Mara Garza advance an argument along these lines, called the *No-Relevant-Difference Argument* (NRDA) (Schwitzgebel & Garza, 2015, 99):

> P1 If A deserves some degree of moral status and B does not deserve the same degree of moral status, then there must be some relevant difference between A and B that grounds the difference in moral status.
> P2 There are possible AIs or social robots who do not differ in such relevant respects from entities (e.g., human beings) who currently possess moral status.
> C1 Therefore, there are possible AIs who deserve some degree of moral status.

Schwitzgebel and Garza do not offer an extended defense of the first premise, but simply insist that its denial would render "ethics implausibly arbitrary" (Schwitzgebel & Garza, 2015, 99).

[8] Proponents of this view include Coeckelbergh (2010, 2014, 2018, 2022a), Gordon (2021, 2022a), and Gunkel (2011, 2014, 2018).

The second premise is general enough to permit disagreement about which properties ground moral status and to allow agnosticism about whether social robots will someday display those properties (given, for instance, reasonable expectations about technological innovation).[9] Perhaps unsurprisingly, optimism that social robots either presently or will soon possess moral status tends to be a function of the properties one takes to be status conferring.

Schwitzgebel and Garza adopt a permissive "psycho-social view of moral status" according to which only psychological properties and social properties are status conferring (Schwitzgebel & Garza, 2015, 100–1).[10] They also defend the second premise against possible objections, noting that the claim is quite modest given its modality. While *some* artificially intelligent entities (e.g., smartphones) appear not to instantiate morally relevant psycho-social properties, it is difficult to argue that *no* artificially intelligent entity will ever instantiate such properties. As these authors observe,

> no *general* argument has been offered against the moral status of all possible artificial entities. AI research might proceed very differently in the future, including perhaps artificially grown biological or semi-biological systems, chaotic systems, evolved systems, artificial brains, and systems that more effectively exploit quantum superposition (104).[11]

While Schwitzgebel and Garza articulate a relatively standard individualist position within debates about the moral status of SRs, others argue for an even more permissive view. For example, Luciano Floridi has advanced a stronger—and much more controversial—framework, *information ethics*, which affords some degree of moral status to anything that can be considered an information object (Floridi, 1999, 2002).[12]

Other moral individualists—call them *robot rights skeptics*—deny that social robots will likely soon (or ever) possess status-conferring properties.[13] These writers accept the first premise of the NRDA while denying the second.

One motivation for skepticism stems from what Adam Andreotta has recently called *the hard problem of AI rights*. Like many other writers, Andreotta takes phenomenal consciousness to be a necessary condition for moral status ascription

[9] Some writers take the issue to be whether social robots will *ever* possess status-conferring properties, whereas others focus more on the question of whether social robots will likely soon possess those properties.

[10] In admitting that social properties are status conferring, these authors are not thereby committing themselves to a relationalist view. As I shall explain below, MR is not (necessarily) the idea that moral status is conferred by virtue of relationships. Rather it consists in both a negative (anti-individualist) dimension as well as a positive dimension, which suggests that moral status ascriptions should be arrived at through various forms of critical reflection.

[11] For a discussion of the likelihood that these technologies will be available in the (relatively) near future, see Bostrom (2014).

[12] For critical discussions of Floridi's information ethics—especially as it bears on the questions of robot rights—see Brey (2008), Coeckelbergh (2010, 217), Gunkel (2014, 122–6) and Mosakas (2021, 436–8).

[13] See for example Andreotta (2021), Mosakas (2021), Müller (2021), and Veliz (2021). For skeptical arguments grounded in the adverse social implications of ascribing moral status to SRs see Turkle (2011) and Bryson (2009). Relatedly, others have argued that, in principle, SRs are incapable of being moral agents (Sparrow 2021).

(Andreotta, 2021, 24).[14] Our moral concern for humans and non-human animals, for example, is justified only if such creatures undergo subjective experiences and are capable of suffering (24). Despite its popularity, this view seems to run up against a serious problem: given the absence of an agreed upon theory of what consciousness *is* or how it arises, how can one justify attributing moral status to *anything*? In other words, it seems difficult to even apply the consciousness criterion of moral status barring a solution to *the hard problem of consciousness* (i.e., the problem of explaining why physical processes generate qualia).[15] When it comes to the moral status of other humans and non-human animals, Andreotta thinks there is a way around this issue. One can attribute conscious experiences to non-human animals because of their behavioral, evolutionary, and biological similarities to us (Andreotta, 2021, 25). But since there are comparatively few (if any) relevant evolutionary or biological similarities between humans and social robots, we cannot appeal to this explanatory strategy. As Andreotta observes,

> Given that advanced AIs will likely be constituted in ways that are very different to us… current approaches to animal consciousness do not map well to questions of AI consciousness. The 'Hard Problem' for AI rights… stems from the fact that we still lack a solution to the 'Hard Problem' of consciousness (Andreotta, 2021, 19).

There is a well-known thought experiment that challenges the idea that evolutionary and biological similarities provide our only justification for attributing phenomenal consciousness to other entities (Schwitzgebel & Garza, 2015, 103; Chalmers, 1995). Consider a phenomenally conscious human whose brain is gradually replaced by silicone chips. If, as seems intuitively plausible, minor replacement would not result in the loss of phenomenal consciousness, then one should not expect a different result were the process to be carried out iteratively until the brain was transformed completely.[16] So long as the patient's behavior remained largely unchanged—and especially if they continued reporting the presence of conscious experiences—there does not seem to be grounds to deny that they would still be conscious.

In addition to this thought experiment, there are other ways for moral individualists to avoid the skeptical conclusion that the hard problem of AI rights entails. On the one hand, one might circumvent the biological-evolutionary explanatory strategy by appealing to empirical tests for consciousness. Recently, for example, Susan Schneider has proposed several frameworks for testing whether a machine is conscious (Schneider, 2019).[17] On the other hand, an individualist could deny Andreotta's

[14] See also Mosakas (2021, 431).

[15] For a discussion, see Chalmers (1995).

[16] Andreotta does not find this line of argument convincing given its reliance on intuitions about cases for which we have no empirical support (Andreotta 2021, 27–8).

[17] The first, "AI Consciousness Test" is meant to serve as a sufficient, but not necessary condition for determining consciousness (Schneider 2019, 50). It attempts to "challenge an AI with a series of increasingly demanding natural language interactions to see how readily it can grasp and use concepts based on the internal experiences we associate with consciousness" (51). For critical discussions of Schneider's tests, see Andreotta (2021).

claim that consciousness is a necessary condition for moral status. As mentioned above, Schwitzgebel and Garza's psycho-social view leaves open the possibility that social properties may be sufficient to ground the moral status of entities that lack conscious experience. And Floridi's information ethics allows that an entity can have moral status without being consciously aware (Floridi, 2002).[18]

2.2 Problems with Moral Individualism

Despite its predominance within debates about the moral status of social robots, MI faces several challenges. In this section, I examine these objections and contend that they suggest the need for a *constrained individualism*, which embraces a limited role for status-conferring properties within moral practice while rejecting the claim that these properties produce agent-neutral reasons. This position will serve as a premise in my argument for a pragmatic approach in Section 3.

Mark Coeckelbergh argues that moral individualism encounters *conceptual* and *epistemological problems*, which make it difficult or impossible to ascribe moral status in practice, and which generate interminable disagreements about the limits of moral concern (Coeckelbergh, 2010). Consider problems involving *semantic indeterminacy*. Virtually every property that the individualist deems "status-conferring" is vague or ambiguous. As David Gunkel puts it, terms such as *sentience*, *consciousness*, *rationality*, or *agency*, are "undecided and considerably equivocal", and have contested meanings not only within philosophy, but across psychology, neuroscience, and robotics (Gunkel, 2014, 116). But without a clear understanding of what these concepts mean, it is hard to see what licenses the moral individualist to use them in ascribing moral status. More generally, the problem of semantic indeterminacy belongs to a wider philosophical issue that encapsulates not only status-conferring properties, but normative concepts as well.

This criticism depends on two more general claims. First, that one cannot be justified in valuing x, without having a clear sense of what x is. And second, that moral individualism cannot be action-guiding unless one can determine which entities instantiate status-conferring properties. The first point has to do with indeterminacy involving a term's intension, whereas the second has to do with its extension. While I think there is some plausibility to these objections, I worry that critics of moral individualism overestimate their decisiveness. An individualist could concede that their theories are predicated on contested concepts, while denying that this poses unsurmountable problems. For example, Andreotta suggests that our intimate familiarity with first-personal conscious experiences licenses our use of the term "consciousness" in moral theorizing (Andreotta, 2021, 25). We may not be able to define or explain the nature of phenomenal experience, but it is something with which each of us is, presumably, well-acquainted.

Semantic inferentialism provides a more promising (and to my knowledge, unexplored) way of developing the problem of semantic indeterminacy. Rather than

[18] Neely argues that it is possible for AI to have interests even if they lack phenomenal consciousness, and that this suffices for their having moral status (Neely 2014).

frame the issue in terms of our concepts' intensions or extensions, the critic might argue that the moral individualist relies on a cluster of concepts whose inferential relationships are poorly understood or obscure. Consider, for example, the relationship between *sentience* and having *interests*. Even if there were greater consensus about the definition or referent of these terms, one might still wonder about the entailment relations between them. Some philosophers contend that sentience is a necessary condition for having interests (DeGrazia, 2008). On this view, someone who claimed to be concerned about their plant's interests (e.g., in being watered) would be exhibiting a conceptual confusion (or speaking metaphorically). Other authors think that these notions can be held apart, such that one can intelligibly attribute interests to nonsentient entities or envision cases in which a sentient being did not have interests (Neely, 2014, 98).[19]

In response, a moral individualist might accept that within ordinary discourse inferential connections between status-conferring concepts are messy, but insist that they can be given greater determinacy when contextualized. It may be pointless to ask *in general* whether having interests entails being sentient. What matters is that within a localized sociolinguistic practice such as a scientific theory or a legal framework, these relationships can be more clearly specified. In other words, it is possible to avoid the problem of inferential semantic indeterminacy by relativizing inferential relations to particular language games. Within a well-defined theoretical context, for example, one can clearly specify the connections between concepts such as *sentience, interests,* or *harm*. But in doing so, one must concede that whatever normative conclusions follow from the arguments employing those concepts are only likely to appear compelling to participants in those practices. This response—as I shall discuss in greater detail below—is made available if one adopts a constrained version of individualism.

Setting aside questions of meaning, another objection to MI is that there is no principled way of determining which intrinsic properties are morally relevant. Call this the *problem of relevance*. As Coeckelbergh puts it, since "[o]ur moral intuitions differ on what criteria are the relevant ones", questions about which properties ground moral status generate unavoidable disagreements (Coeckelbergh, 2010, 212). Many people will readily concede, for example, that a dog's capacity to feel pain serves as a perfectly good reason not to kick him. Indeed, many moral individualists predicate their claim that sentience is a sufficient condition for possessing moral status on the intuitive idea that causing pain to a sentient being is inherently morally objectionable. For critics of MI, however, it is always possible to envision someone who does not share this intuition. There are people whose background beliefs and desires are configured in such a way that they do *not* take causing pain to a sentient being to be inherently objectionable. A religious fundamentalist, for example, might simply regard as irrelevant the capacity to experience pleasure or pain. When it comes to attributing moral status, they might insist that the only thing that *really*

[19] Another example of the inferential indeterminacy would be the question of whether *consciousness* is conceptually separable from notions such as *intelligence* or *rationality*. Andreotta argues that these notions are independent of one another, such that it is possible to have an intelligent machine that is not phenomenally conscious (Andreotta 2021, 22–23). But one could envision someone who denied this claim of independence.

matters morally is the possession of a soul. Of course, moral individualists who take sentience to be a morally relevant property will claim that the would-be fundamentalist is being unreasonable. But it is unclear what force this sort of consideration is supposed to have.

For critics of MI, what motivates the problem of relevance is skepticism toward the claim that any property can have intrinsic moral relevance (and therefore, that it can provide agent-neutral reasons for action). On their view, relevance is a relative notion. While it may be true that *many* people take sentience to be a morally relevant property (and therefore, view it as a reasonable condition for ascribing moral status), critics of MI take this to be a contingent fact about those people and not a necessary feature of sentience.

Some critics of MI, whose views I shall discuss in the next section, take the problem of relevance to necessitate a radically different approach to moral status (Coeckelbergh, 2010; Coeckelbergh & Gunkel, 2014; Gunkel, 2014). That is, they take the problem (in conjunction with the other issues discussed above) to undermine the very idea that *properties* can ground moral status. Ultimately, I shall I contend that these conclusions are too hasty. The problems discussed in this section are best addressed not through a total abandonment of moral individualism but by amending it.

One way of circumventing the problem of relevance is to adopt a *constrained individualism,* which holds that the question of which properties are morally relevant can be contextualized. Adapting a central insight of metaethical constructivism, the constrained moral individualist maintains that whether an agent has reason to regard a certain property as status conferring is a function of their background practical identities (Korsgaard, 1996, 2009; Street, 2012).[20] As Christine Korsgaard puts it, "[o]ur conceptions of our practical identity govern our choice of actions, for to value yourself in a certain role or under a certain description is at the same time to find it worthwhile to do certain acts for the sake of certain ends, and impossible, even unthinkable, to do others" (2009, 20). Put simply, a property's moral relevance is a contextual notion, which depends on the background beliefs and values constitutive of one's practical identity. This perspective preserves the individualist's idea that a moral patient's properties can factor into our reason-giving practices, while abandoning their claim that these reasons can be agent-neutral. This proposal will ultimately play an important role in my argument for a pragmatic view of moral status, which I shall discuss in detail below.

2.3 Moral Relationalism and Social Robots

In this section, I consider a family of approaches to the moral status of SRs that typically sets itself in opposition to moral individualism. This view, which I have been referring to as *moral relationalism*, rejects the idea that an entity's moral status

[20] This idea is, however, not limited to metaethical constructivism, but has been developed in considerable detail within other areas of philosophy—notably feminist philosophy (Lindemann 2019, chapter 4) and pragmatism (Rorty 1989). It finds empirical support from social identity theory and self-categorization theory (Jenkins 2014).

depends on its properties. In doing so, MR hopes to reorient debates about the moral status of SRs toward alternative metatheoretical commitments. Like MI, MR is not a univocal position, but includes a constellation of philosophical views sharing common features. While this discussion draws primarily from the relational accounts of moral status developed by Mark Coeckelbergh and David J. Gunkel, theirs are by no means the only such approaches. A growing number of scholars have developed relationalist approaches that draw from non-Western cultural perspectives—especially African conceptions of personhood (Coeckelbergh, 2022b; Jecker et al., 2022a, b; Wareham, 2021).

Moral relationalists contend that entities cannot be adequately understood or defined apart from the social and natural relations in which they stand to other entities (Coeckelbergh, 2014, 64). This is not a metaphysical claim that an entity's moral status is grounded in its relations to other entities, but rather a methodological claim about how inquiry about an entity's moral status ought to proceed. As Mark Coeckelbergh explains, to substitute a relational ontology for a properties-based one would be tantamount to substituting one "dogmatic" approach for another (Coeckelbergh, 2014, 65). The relationalist's contention is that moral status is always ascribed within a complex socio-historical context, and that this context ought to be the primary site of critical moral reflection. When it comes to the moral status of SRs, adopting methodological relationalism requires one to interrogate their "relations with other machines and with humans" and to understand how these entities are "naturally, materially, socially, and culturally embedded and constituted" (64).

This commitment to methodological relationalism is closely connected to a second feature of MR, namely that moral agents and moral patients are mutually co-constituting. Relationalists typically put this point as a rejection of the Cartesian subject-object dichotomy. On this view, moral status is not something that exists antecedently to and independently of social relations (or of any relations, for that matter). Rather, it is "itself the outcome of the process of relation and interaction" (Coeckelbergh, 2018, 149). This is not to say that moral status ascription is simply a matter of fiat. Rather, we find ourselves "thrown" into a world of extant social relations, within which the limits of moral concern appear natural or fixed. As Coeckelbergh explains,

> when I, as a moral subject, "ascribe" moral status to an entity, I am not the first one to do so and the way I do it and the status I ascribe are probably already available in my society, my culture, and my language – more generally in what Wittgenstein would call my 'form of life'… Therefore, the question of moral standing is always connected to the question who is part of the moral *community* and what moral games are already played when and before I ask the question (149).

A third commitment underlying moral relationalism concerns the target of moral reflection and deliberation. The fact that moral status ascription is always the outcome of context-dependent social relations suggests the need for a *critical* or *transcendental perspective*. Rather than inquire into the grounds of moral status, a relational view interrogates the conditions of possibility for such ascription (Coeckelbergh, 2014, 64). That is, it implies that "we need to reveal and criticize the social

background of the question" (Coeckelbergh, 2018 149). Consider, for example, how an entity's moral status is "partly constituted by the way we talk about it" (Coeckelbergh & Gunkel, 2014, 724). In particular, Coeckelbergh and Gunkel claim that our practices of *naming* have significant moral consequences, insofar as they function as a "way of demarcating the moral community" (725). For example, that people seldom consume their pets—plausibly—has to do with how they relate to them. One way in which these relations are expressed is through "bestow[ing] singular proper names on an individual, and thereby individuated, animal" (725). As Coeckelbergh and Gunkel observe, "[w]e call this specific dog 'Lassie' or that cat 'Mister Wiskers.' When an animal is named in this fashion, it often takes on face and is protected from abuse and killing. It becomes a 'pet', a 'family member', etc. rather than an 'animal'" (725).[21] From this perspective, moral individualism misses an important site for critical moral reflection and transformation. Naming practices are an important feature of our moral landscape that ought to be subject to moral reflection. But names are not properties. They are relational phenomena that cannot be adequately understood apart from the complex social interactions in which they are embedded.

As Coeckelbergh and Gunkel acknowledge, compared to moral individualism, MR offers little by way of definite, prescriptive recommendations (Coeckelbergh & Gunkel, 2014, 728; Coeckelbergh, 2018, 153). As they observe, "[t]his analysis of conditions of possibility for relations does not in itself advance a straightforward normative position" (728). Unlike the moral individualist, relationalists do not, for instance, offer a set of necessary or sufficient conditions for attributing moral status. Their approach leaves things at a greater level of indeterminacy, which they maintain, is unavoidable.

Nonetheless, one can glean some positive recommendations from their proposals. First, relationalism motivates the cultivation of a set of meta-normative or meta-critical attitudes. As Coeckelbergh puts it, "a cautious, patient, and open attitude (and indeed character), then, can be said to constitute a meta-moral demand and a meta-virtue or moral-epistemic virtue" (Coeckelbergh, 2018, 156). Given that moral status attributions are the outcome of relations and interactions that are themselves evolving, one ought to acknowledge that one's own attitudes and commitments concerning the limits of moral concern are likely to evolve as well.[22]

A second positive feature of moral relationalism is that we ought to take seriously the role of art in moral edification and reflection. Considering a number of performance pieces and installations querying the boundaries between humans and machines, Coeckelbergh writes:

> works of art such as these invite us to destabilize and critically question established meanings and borders, here to question the sharp border between machines and humans, or at least invite us to consider how in our imagination and feeling we already easily cross this border – whatever science or metaphysics may tell us (Coeckelbergh, 2018, 155).

[21] The notion of "taking on face" is one that Gunkel and Coeckelbergh derive from Emmanuel Levinas.

[22] In response, scholars have recently argued that, despite its own claims to promote a critical or reflexive attitude, MR ends up perpetuating anthropocentric biases (Gordon 2022b; Sætra 2021).

Finally, a relational approach requires that one take seriously "the phenomenology and experience of other entities such as robots" (Coeckelbergh, 2018, 149), by interrogating how these entities appear to us through our embodied social relations with them (2018, 149; 2010, 214; 2014, 64). This requires contending with our emotional and affective responses to social robots, rather than viewing these phenomena as "mere appearances" or mistakes in need of correction.

2.4 Problems with Moral Relationalism

Moral relationalist attempts to rethink the moral status of social robots beyond the individualist paradigm have encountered considerable resistance. The most common objection is that relationalism amounts to an untenable form of relativism. That is, "taking the relational turn," renders impossible rational disagreement about moral status ascription, and leaves us with little to no normative guidance. This objection is expressed by Kestutis Mosakas, who writes:

> It seems that what the relational approach is fundamentally concerned with is our feelings and attitudes towards different entities, since that is what constitutes the basis of our relations; but without any central moral properties or guiding principles, it is difficult to see how this approach could genuinely help us in our moral decision-making without getting bogged down in a sea of relative judgements (Mosakas, 2021, 434).

Similarly, Vincent Müller objects that "the core of the relational turn" is simply a "version of anything goes that dissolves the question of moral patiency to a random act of will" (Müller, 2021). An unacceptable implication of MR is that "anything I happen to care about receives moral status" (Müller, 2021). Thus, for relationalists, no manner of ascribing moral status can be considered "better" or "worse" than any other.

While there is certainly merit to these concerns, I worry that individualist critics tend to overstate their case when they claim that moral relationalism entails an "anything goes" approach to moral disagreement. As we have seen, relationalists do provide limited forms of normative guidance by recommending meta-normative values—such as open-mindedness—that ought to govern moral deliberation (Coeckelbergh, 2018). These values can serve as a basis for rational critique of extant moral status ascriptions, and can go some way towards adjudicating between competing views about the limits of moral concern. For example, a relationalist could maintain that one mode of ascription is preferable to another on the basis of the fact that its proponents have more thoroughly considered the issue. More importantly, nowhere, to my knowledge, has a moral relationalist ever *denied* that agents ought to offer *reasons* in support of their judgments about moral status. What they deny is that one can legitimately derive agent-neutral normative reasons by inspecting an entity's non-moral properties.

A second objection is that relational approaches fail to do justice to our considered moral judgments about particular cases (Mosakas, 2021, 436). Critics worry

that MR ends up denying moral status to entities who *ought* to be afforded moral status when those entities fall outside of appropriate social relations. Consider, for example, what Mosakas calls the *Robinson Crusoe Problem*: if relationalism is true, then Robinson Crusoe (i.e., a person stranded on an island, standing in no social relations to others) would lack moral status, whereas Paro (i.e., a non-sentient baby seal-shaped robot used to treat dementia patients) would be morally considerable. For Mosakas, "that is a problem, because, in case of an ethical dilemma, it would seem that no one in their right mind should morally prioritize Paro over Crusoe" (435). Any theory that so flagrantly violates our moral intuitions ought to be rejected.

In response, a moral relationalist might concede that while a Robinson Crusoe figure would lack moral status in theory, it is difficult to conceive of a practical context in which such a figure could ever exist. Given that a human infant would not survive without the care of others, it is unclear how Crusoe could come to exist without standing in social relations to others (arguably, from the time of conception, all humans stand in social relations by virtue of existing with another human body, or through connections to their biological parents and those who nurture them). Perhaps the idea is that Crusoe stands in no *present* relations to others—such that anyone to whom they were once related has either permanently forgotten this fact or died (presumably, there would have to be no known records of Crusoe, since such knowledge would arguably generate some sort of social relation). But even if such extraordinary circumstances were to occur, what gives Mosakas's argument its force is the possibility of an actual moral dilemma in which *somebody* must decide between prioritizing Crusoe or Paro (Mosakas, 2021, 435). At this point, a moral relationalist could insist that by virtue of having to make such a decision, the would-be moral decision-maker would thereby enter into a social relation with Crusoe. Thus, by the relationalist's standards Crusoe would have moral status.

Nonetheless, I agree that there is a legitimate worry about the *degree* of normative guidance that relationalism is able to offer. One reason for dissatisfaction with this position is that we do ultimately encounter practical contexts in which there are unavoidable disagreements about the moral status of SRs. Even someone who rejects the existence of exceptionless moral principles could rightly demand *some* form of normative guidance that extends beyond what the relationalist has provided.

In addition to these concerns about normative guidance, I submit that there is a more serious problem with relational approaches, namely, that they lack a *theory of error* which can explain why properties-based approaches seem to capture so many people's moral intuitions. Arguably, many people *do* appeal to an entity's properties when justifying how it should be treated. For many moral agents, it is plausible that our moral concern for others is grounded in (or at least reflects our sensitivities to) their properties or attributes. *Because she can feel pain* is, on the face of it, a reasonable response to the question of why one ought not to pull the cat's tail. Although proponents of the relational turn raise compelling conceptual, epistemic, and practical problems for a *general theory* of moral status limited exclusively to intrinsic

properties, they fail to explain moral individualism's intuitive appeal and its evident role in moral-discursive practice.[23]

The relationalist must either explain these intuitions away or find a way to accommodate them. The proposal developed in this paper opts for the latter option. Relationalist approaches go too far in completely eschewing status-conferring properties from moral practice; and, consequently they fail to appreciate the possibility of a more narrowly circumscribed version of the property-based view. The *constrained individualism*, which I introduced above, affords properties a role within moral reason-giving, while denying that they can be understood independently of their relations to an agent's background beliefs and values. Similarly, one might adopt a *constrained relationalism*, according to which, transcendental social critique, projects of imaginative expansion, and other deliberative strategies proposed by relationalists are indispensable for *some forms* of inquiry into our treatment of SRs, while conceding that properties can be morally relevant within other deliberative contexts. I now turn to a defense of this position.[24]

3 Toward a Pragmatic Account of Moral Status

I have been arguing that the challenges to moral individualism and moral relationalism push both views in the direction of more constrained positions. Moral individualism can confront the problems of semantic indeterminacy and relevance by jettisoning the claim that status-conferring properties provide agent-neutral reasons for action. This can be achieved, I suggested, by relativizing the relevance of status-conferring properties to background practical identities. Likewise, a constrained form of moral relationalism will concede that the identification of status conferring properties can play *some* role in moral reason-giving practices. In what follows I shall operate with constrained versions of both MI and MR in mind.

[23] In claiming that relationalists require a theory of error, I do not mean to claim that appealing to status-conferring properties to ground moral status is a universal practice. Indeed, traditional sub-Saharan African and contemporary Japanese societies do not rely on individualist intuitions (Jecker and Nakazawa 2022). My claim is that insofar as relationalists deny the legitimacy of individualist justifications, they owe an explanation not only of *why* these justifications are mistaken, but of *how* such a justificatory error became so prevalent—especially within post-enlightenment Western societies.

[24] Constrained relationalism bears important similarities to other positions within the theoretical landscape. John Danaher has recently advanced a view called ethical behaviourism (EB), according to which an entity has moral status if it consistently behaves like other entities to which we ascribe moral status (Danaher 2020). Both EB and constrained relationalism are compatible with a range of views about how moral status attributions are justified (Danaher 2020, 2024). Moreover, both views prioritize normative and epistemological questions about moral status over metaphysical ones (Danaher 2020, 2027). But whereas EB focuses on justifying empirical inquiry into an entity's behavior as a primary means of determining its moral status, constrained relationalism focuses on a broader range of deliberative strategies, including transcendental critique, sociolinguisitic analysis, phenomenological inquiry, and so on. Constrained relationalism also shares important affinities with pluriversal approaches to ethical thought (Reiter 2018). While a detailed comparison is beyond the scope of this paper, both positions are amenable to multiple legitimate methods and forms of inquiry in ethics. They are also resistant to the assumption that all ethical questions—including those about the limits of moral considerability—admit of a single answer that holds universally.

While this compromise goes some way toward reconciling MI and MR, more needs to be said about how the two positions can be integrated or synthesized. Recently, John-Stewart Gordon and David J. Gunkel have suggested that both MI and MR may be needed to face the ethical challenges that AI and social robots pose (Gordon, 2021; Gordon & Gunkel, 2022). While I agree with the direction of their proposal, it leaves open questions about how the two views can be rendered compatible. In particular, a key challenge that remains is to explain what synthesizing MI and MR would look like in practice, especially given cases where they are likely to conflict. The pragmatic approach that I shall defend aims to throw light on these questions.

The main argument presented in this section is a pragmatic one. The key to integrating constrained forms of MI and MR is to consider them as distinct theoretical toolkits, or *deliberative strategies* that answer to different problems that agents face. On the one hand, I shall argue that MI offers a set of *value simplifying* strategies that are especially useful in addressing moral coordination problems (i.e., problems in which agents with potentially different moral outlooks must agree on how to act). On the other hand, I maintain that MR offers a set of strategies for preserving the complexity of our values, which are indispensable for moral edification and social criticism.

This proposal requires a radical—but on my view, fruitful—shift in perspective. Rather than viewing MI and MR as competing theories about the *nature* of moral status, I recommend that philosophers turn their attention to neglected *pragmatic* questions about the point or function of MI and MR. The advantages of this shift in focus are twofold: First, it lends further support to the claim that constrained versions of MI and MR are compatible. Instead of theoretical rivals, both perspectives are construed as tools that answer to different sets of problems. Second, shifting philosophical attention to moral individualism's and relationalism's functions offers a blueprint for how they can be integrated in practice. That is, a pragmatic approach can help determine when MI ought to be prioritized over MR and vice versa.

I shall outline this pragmatic approach in several steps. In Section 3.1 and 3.2 I explain what it means to construe MI and MR as *deliberative strategies.* In doing so, I shall highlight the central functional advantages of each strategy and discuss the practical contexts for which they are best suited. Next, Section 3.3 summarizes key advantages of a pragmatic approach. Finally, Section 3.4 offers several broad-strokes suggestions of what such an approach might look like in practice.

3.1 Moral Individualism and Value Clarity

When it comes to determining which entities have moral status and how those entities ought to be treated, moral individualists offer the following deliberative strategy: identify a set of properties that are status-conferring, and then determine which individuals instantiate those properties. A pragmatic approach to moral status attribution invites several questions about this deliberative strategy: What is the point of applying it? In which contexts is such a strategy called for?

I submit that MI offers a deliberative strategy that discharges two central functions. On the one hand, it affords agents a kind of *value clarity* about the scope of

their moral obligations. On the other hand, it aims to resolve moral coordination problems amongst agents with potentially different evaluative standpoints. Allow me to discuss both functions in turn.

So often, our moral outlooks involve a welter of inchoate ideals and values. Although they may seem natural or intuitive (from the inside, so to speak) it can often be difficult to explain or justify our moral perspectives. In particular, it can be especially difficult to justify why we feel compelled to extend our moral consideration towards certain moral patients rather than others. By distilling questions of moral status to a set of objective properties, MI offers agents a significant source of value clarity that helps them articulate and justify their moral views.

Consider, for example, the normative complexity and practical indeterminacy that a person is likely to experience when adopting an ethical commitment to environmentalism. On the one hand, an environmentalist may find certain courses of action to be required of them, without being able to articulate precisely why. They may, for instance, feel a sense of direct obligation to preserve *ecosystems* even without being able to specify clearly their reasons for doing so. Perhaps they might appeal to their love of nature, the importance of conservation, or some desire to promote biodiversity—perhaps they may simply appeal to the fact that they care about the environment. On the other hand, their commitment to environmentalism may generate tensions and conflicts with their other moral commitments and intuitions. They may find themselves confronted with questions about the adverse environmental impacts of their profession or lifestyle. Perhaps they may find their commitment to environmentalism to be at odds with their commitment to animal rights (Sagoff, 1984).

Moral individualism offers a helpful deliberative strategy in such cases. It allows an agent to effectively cut through this morass of values and sentiments, thereby delivering a kind of *value clarity* about the scope of her obligations and the reasons underlying them. Many environmentalists, for example, have been attracted to *biocentrism*—a framework that enables them to maintain that all living things have intrinsic value. This form of moral individualism serves to simplify and clarify the complex values and motivations underlying an environmentalist moral outlook in order to facilitate deliberation and the justification of their projects.

In addition to providing value clarity, MI can serve as a powerful mechanism for solving moral coordination problems. By distilling questions of the basis of moral status down to a simplified set of properties, MI can facilitate rational discussion about the limits of moral concern, especially when the parties to such disagreement have diverging moral outlooks. This mechanism works by identifying some set of properties that are likely to appear morally salient from as wide a range of perspectives as possible. In doing so, it can motivate moral rules or policies that aspire to a high degree of generality and scope.

Consider, for example, a group of researchers charged with developing policies governing the treatment of research subjects (human or non-human). Given the possibility of participants whose moral intuitions, values, attitudes, and beliefs may vary considerably, the policymakers face a coordination problem: they must design rules for the treatment of research subjects that will appear reasonable from a potentially wide range of moral perspectives. Here, MI's *value clarifying* strategy is especially fruitful. It enjoins the policymakers to identify the kinds of properties that are likely

to seem morally salient to a wide variety of potential participants and to develop policies that focus on the treatment of entities who instantiate those properties.

It is no surprise, for example, that sentience plays a key role in the deliberations of research ethics boards charged with assessing experiments on non-human animals, or that patient autonomy factors so significantly in the field of medical ethics. Scientific research and clinical medicine are practical contexts in which we are likely to face moral coordination problems. They involve institutions in which moral agents with varying normative outlooks must all agree on how to behave. Both sentience and autonomy are excellent candidates for status-conferring properties in these contexts because they are likely to appear morally relevant from a wide range of perspectives.

3.2 Moral Relationalism: Preserving Value Complexity

On the face of it, MR may appear to offer a deliberative strategy that closely resembles that of MI. One might expect it to operate as follows: first identify which social relations are relevant, and then determine whether an entity stands in the appropriate relationships to others. On my view, however, this characterization is misleading. To put the point in the terminology I introduce above, to see MR as recommending this strategy is to construe it in metaphysical rather than methodological terms—a characterization that many relationalists outright reject (Coeckelbergh, 2014). In claiming that an entity's moral status is intertwined with its social relations, the relationalist is *not* attempting to state the necessary and sufficient conditions for something's having moral status (Coeckelbergh, 2014; Jecker et al., 2022b). Rather, the purpose of such a claim is to reorient the focus of moral inquiry toward the conditions under which moral status is ascribed. While different relationalist accounts emphasize different sets of strategies for undertaking such inquiry, they all share a sensitivity to the complexity and indeterminacy of those moral contexts in which moral status is ascribed.

Allow me to reiterate some of the various relationalist *complexity preserving* deliberative strategies mentioned above in (Section 2.3). Moral relationalists recommend:

(i) *Transcendental inquiry* into the historical, social, political, and economic conditions under which the limits of moral community are drawn (Coeckelbergh, 2010)
(ii) *Sociolinguistic inquiry* into how our linguistic practices (e.g., naming) shape the boundaries of our moral communities (Coeckelbergh and Gunkel, 2014).
(iii) *Phenomenological inquiry* into the affective or experiential dimensions of our relations to entities (Gunkel, 2014, 2018).
(iv) *Social-relational inquiry* into the normative quality of our relations to other entities or their roles within the moral community (Jecker et al., 2022a).

In contrast to MI—which aims to clarify and simplify our values—these complexity preserving strategies are not geared toward securing moral agreement with others. Rather, they are indispensable mechanisms of *moral edification* and *social criticism*.

By engaging the imagination and subjecting our moral intuitions to critical scrutiny, MR can facilitate forms of moral transformation that MI struggles to provide.

For example, while MI may help clarify and justify our values, it is unlikely to lead anyone to transform their intuitions about which properties confer moral status in the first place. By contrast, while moral relationalism may not offer tools for solving moral disagreements, it promotes the sorts of inquiry that *can* lead people to dramatically revise their moral intuitions. Complexity preserving deliberation can lead us to regard as morally relevant features that were once seen as unimportant.

3.3 Advantages of a Pragmatic Approach

I have argued that MI and MR recommend deliberative strategies that are useful within different practical contexts. Moral individualism offers *value simplifying* strategies that are best suited for contexts in which we need to clarify our values or to coordinate our behavior with others. Moral relationalism offers complexity preserving strategies that have the potential to challenge and transform our moral outlooks. By engaging the imagination and critiquing our intuitions these strategies are invaluable sources of moral edification and social criticism. Having outlined this pragmatic approach, allow me to state several of its advantages.

First, this approach allows one to better appreciate the compatibility of MI and MR. So long as the two are treated as competing theories about the nature of moral status, they are bound to appear irreconcilable. By focusing instead on the practical effects of adopting individualist or relationalist standpoints—that is, by focusing on the deliberative strategies they offer—it is possible to view them as answering to distinct practical needs. Their complementarity lies in their being different tools for different purposes.

A second advantage of this approach is that it can help make sense of cultural variability in moral status ascription. Recent studies have suggested significant differences in cultural values expressed towards technology in general and social robots in particular (Jecker & Nakazawa, 2022). When it comes to how moral status is understood and justified, a growing number of authors have argued that African conceptions of personhood embody a relational conception of moral personhood rather than an individualist conception (Coeckelbergh, 2022b; Jecker et al., 2022a, b; Wareham, 2021).

The pragmatic approach developed in this paper is well positioned to account for this cultural variability. Rather than assume that some cultures have privileged insight into the nature of moral status that others lack, the relative presence of individualist or relationalist tendencies across cultures may be explained by differences in the kinds of practical problems that these cultures face. One hypothesis (that merits further exploration) is that cultures that experience a greater prevalence of the kinds of moral coordination problems described in Section 3.1, will be more reliant on moral individualism and its associated model of personhood.

Finally, a pragmatic approach offers a blueprint for deploying both sets of strategies in practice. It helps us better understand when to prioritize MI or MR. Generally speaking, I have argued that the former is especially useful in contexts in which

we face moral coordination problems, whereas the latter is better suited toward contexts in which we are aiming for moral edification and cultural criticism. In what follows, I develop these suggestions in greater detail.

3.4 Applying the Pragmatic Approach

One implication of the pragmatic approach is that it advises against speaking about the moral status of social robots *in general*. Rather, than begin by investigating the *nature* or *grounds* of moral status, on a pragmatic approach one would need to begin by identifying the practical contexts in which moral individualism and moral relationalism's deliberative strategies are called for. I have already offered a general characterization of how this might look. But allow me to spell this out in more detail.

First, one feature of the constrained form of individualism I recommend is its *domain specificity*. Rather than ask which properties ground moral status in general, one should begin by looking to the contexts where the individualist's value-simplification strategy would be most appropriate. On my view, the most salient contexts are those in which the following two conditions hold:

(i) Moral agents with potentially radically different moral outlooks will need to interact with social robots.
(ii) These situations require agents to coordinate their attitudes concerning the treatment of SRs.

One could envision these conditions being met as SRs are integrated into domains such as healthcare, education, industry, entertainment, and the military. These are domains in which it is necessary to coordinate the behavior of agents whose background beliefs and values often exhibit significant divergence. In these cases, a constrained form of individualism recommends identifying the stakeholders involved in the situation, considering which properties are morally salient from the perspective of those relevant stakeholders, and then developing rules and policies that take those properties to be status-conferring.

By contrast, moral relationalism is most applicable within contexts in which agents are striving for moral self-transformation and social criticism. As social robots increasingly come to inhabit our shared social world, a constrained version of moral relationalism will be especially useful for facilitating reflection on our shared practical identities with those machines. For the most part, the identities and social roles we currently adopt—be it our familial roles, professions, memberships in various organizations, religious affiliations, and so on—are ones we share with other humans. But as social robots develop complex capacities that enable them to *participate* in social practices, it is easy to imagine cases in which we would begin to ask whether *they* could be said to occupy these shared identities as well. The integration of SRs into medical practices would almost certainly raise questions about what it means to be a healthcare provider. Might practitioners someday regard the intelligent machines with whom they increasingly interact and cooperate as "fellow surgeons"?

On my view, these questions are not ones that are best answered by attempting to *simplify* our values in the service of cooperation with others. Rather, these kinds of question require critical and imaginative reflection—that is to say, they demand the sort of complexity-preserving strategies offered by MR.

On a constrained version of MR, an important resource for critical reflection on the possibility of shared practical identities with social robots would be the production and enjoyment of art, literature, and film. These mediums challenge us to rethink our existing practical identities but also to imagine possible future shared identities. In doing so, they can help us reassess the meaning and relevance of the properties and relations we *do* currently see as salient and important. For example, films like *Her*, or *Ex Machina* challenge us to rethink the meaning of notions such as *intelligence*, *friendship*, *suffering*, *agency*, and *trust* through their depictions of human–machine interaction. Art and literature can extend the use of our concepts to new situations, thereby affecting a kind of moral reorientation. These projects of moral self-transformation and social criticism that MR encourages may ultimately lead to revisions in the individualist-oriented policymaking within our shared institutions by leading us to rethink which properties are, in fact, the morally relevant ones.

Acknowledgements I am grateful to Colin Koopman for providing feedback on an earlier draft of this paper. I would also like to thank two anonymous reviewers for their exceptionally helpful comments.

Author Contributions N/a.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Ethics Approval N/A.

Consent to Participate N/A.

Consent to Publish N/A.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Andreotta, A. J. (2021). The hard problem of AI rights. *AI & Society, 36*, 19–32.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 316–334). Cambridge University Press.

Brey, P. (2008). Do we have moral duties towards information objects? *Ethics and Information Technology, 10*, 109–114.

Bryson, J. J. (2009). Robots should be slaves. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues*. John Benjamins Publishing Company.

Cappuccio, M. L., Peeters, A., & McDonald, W. (2019). Sympathy for Dolores: Moral consideration for robots based on virtue and recognition. *Philosophy & Technology, 33*(1), 9–31.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.
Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology, 12*, 209–221.
Coeckelbergh, M. (2012). *Growing moral relations: Critique of moral status ascription*. Palgrave Macmillan.
Coeckelbergh, M. (2014). The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics. *Philosophy & Technology, 27*(1), 61–77.
Coeckelbergh, M. (2018). Why care about robots? Empathy, moral standing, and the language of suffering. *Kairos. Journal of Philosophy & Science, 20*(1), 141–158.
Coeckelbergh, M. (2022a). *Robot ethics*. MIT Press.
Coeckelbergh, M. (2022b). The Ubuntu robot: Towards a relational conceptual framework for intercultural robotics. *Science and Engineering Ethics, 28*, 16.
Coeckelbergh, M., & Gunkel, D. J. (2014). Facing animals: A relational, other-oriented approach to moral standing. *Journal of Agricultural and Environmental Ethics, 27*, 715–733.
Danaher, J. (2017). Should we be thinking about sex robots? In J. Danaher & N. McArthur (Eds.), *Robot sex: Social and ethical implications*. MIT Press.
Danaher, J. (2019). The rise of the robots and the crisis of moral patiency. *AI & Society, 34*, 129–136.
Danaher, J. (2020). Welcoming robots into the moral circle: A defence of Ethical behaviourism. *Science and Engineering Ethics, 26*, 2023–2049.
Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In R. Calo, A. Michael Froomkin, & I. Kerr (Eds.), *Robot law* (pp. 213–231). Edward Elgar.
Darling, K. (2021). *The new breed: What our history with animals reveals about our future with robots*. Henry Holt & Company.
DeGrazia, D. (2008). Moral status as a matter of degree? *The Southern Journal of Philosophy, 46*(2), 181–198.
DiSilvestro, R. (2010). *Human capacities and moral status*. Springer.
Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology, 1*, 33–52.
Floridi, L. (2002). On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology, 4*, 287–304.
Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. Basic Books.
Frank, L., & Nyholm, S. (2017). Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable? *Artificial Intelligence and Law, 25*, 305–323.
Gordon, J.-S. (2021). Artificial moral and legal personhood. *AI & Society, 36*, 457–471.
Gordon, J.-S. (2022a). Are superintelligent robots entitled to human rights? *Ratio, 35*, 181–193.
Gordon, J.-S. (2022b). The African relational account of social robots: A step back? *Philosophy & Technology, 35*, 49.
Gordon, J.-S., & Gunkel, D. J. (2022). Moral status and intelligent robots. *The Southern Journal of Philosophy, 60*(1), 88–117.
Gunkel, D. J. (2011). *The machine question*. MIT Press.
Gunkel, D. J. (2014). A vindication of the rights of machines. *Philosophy & Technology, 27*, 113–132.
Gunkel, D. J. (2018). The other question: Can and should robots have rights? *Ethics and Information Technology, 20*, 87–99.
Harman, E. (2003). The potentiality problem. *Philosophical Studies, 114*, 173–198.
Jaworska, A., & Tannenbaum, J. (2013). The grounds of moral status. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 ed.) https://plato.stanford.edu/archives/spr2018/entries/grounds-moral-status/. Accessed 1 June 2020
Jecker, N. S. (2021a). Nothing to be ashamed of: Sex robots for older adults with disabilities. *Journal of Medical Ethics, 47*, 26–32.
Jecker, N. S. (2021b). You've got a friend in me: Sociable robots for older adults in an age of global pandemics. *Ethics and Information Technology, 23*(Suppl 1), S35–S43.
Jecker, N. S., & Nakazawa, E. (2022). "Bridging east-west differences in ethics guidance for AI and robotics. *AI, 3*(3), 764–777.
Jecker, N. S., Atiure, C. A., & Ajei, M. O. (2022a). The moral standing of social robots: Untapped insights from Africa. *Philosophy & Technology, 35*(2), 1–22.

Jecker, N. S., Atiure, C. A., & Ajei, M. O. (2022b). Two steps forward: An African relational account of moral standing. *Philosophy & Technology, 35*(2), 38.
Jenkins, R. (2014). *Social identity* (4th ed.). Routledge.
Kamm, F. M. (2007). *Intricate ethics: Rights, responsibilities, and permissible harm*. Oxford University Press.
Kenny, A. (1973). *Wittgenstein*. Harvard University Press.
Kittay, E. F. (2005). At the margins of moral personhood. *Ethics, 116*(1), 100–131.
Korsgaard, C. (1996). *The sources of normativity*. Cambridge University Press.
Korsgaard, C. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford University Press.
Lindemann, H. (2019). *An invitation to feminist ethics* (2nd ed.). Oxford University Press.
Marti, P. (2010). Robot companions. *Interaction Studies, 11*(2), 220–226.
McArthur, N. (2017). The case for sexbots. In J. Danaher & N. McArthur (Eds.), *Robot sex: Social and ethical implications*. MIT Press.
McMahan, J. (2005). Our fellow creatures. *The Journal of Ethics, 9*(3/4), 353–380.
Mosakas, K. (2021). On the moral status of social robots: considering the consciousness criterion. *AI & Society, 36*, 429–443.
Müller, V. C. (2021). Is it time for robot rights? moral status in artificial entities. *Ethics and Information Technology, 23*, 579–587.
Neely, E. L. (2014). Machines and the moral community. *Philosophy & Technology, 27*(1), 97–111.
Nørskov, M. (2016). *Social Robots: Boundaries*. Routledge.
Parviainen, J., & Coeckelbergh, M. (2021). The political choreography of the Sophia robot: Beyond robot rights and citizenship to political performances for the social robotics market. *AI & Society, 36*, 715–724.
Rachels, J. (2005). Drawing lines. In C. R. Sunstein & M. C. Nussbaum (Eds.), *Animal rights: Current debates and new directions*. Oxford University Press.
Reiter, B. (2018). Introduction. In B. Reiter (Ed.), *Constructing the pluriverse: The geopolitics of knowledge*. Duke University Press.
Rorty, R. (1989). *Contingency, irony, and solidarity*. Cambridge University Press.
Sætra, H. S. (2021). Challenging the neo-anthropocentric relational approach to robot rights. *Frontiers in Robotics and AI, 8*, 744426.
Sagoff, M. (1984). Animal liberation and environmental ethics: Bad marriage, quick divorce. *Osgoode Hall Law Journal, 22*(2), 297–307.
Schneider, S. (2019). *Artificial you: AI and the future of your mind*. Princeton University Press.
Schwitzgebel, E., & Garza, M. (2015). A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy, 39*(1), 98–119.
Sharkey, A., & Sharkey, N. (2010). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology, 14*, 27–40.
Sparrow, R. (2021). Why machines cannot be moral. *AI & Society, 36*, 685–693.
Street, S. (2012). Coming to terms with contingency: Humean constructivism about practical reason. In J. Lenman & Y. Shemmer (Eds.), *Constructivism in Practical Philosophy*. Oxford University Press.
Tavani, H. T. (2018). Can social robots qualify for moral consideration? Reframing the question about robot rights. *Information, 9*(4), 73.
Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
Véliz, C. (2021). Moral zombies: Why algorithms are not moral agents. *AI & Society, 36*, 487–497.
Wareham, C. S. (2021). Artificial intelligence and African conceptions of personhood. *Ethics and Information Technology, 23*(2), 127–136.
Warren, M. A. (1997). *Moral status: Obligations to persons and other living things*. Oxford University Press.