AI & SOCIETY (2024) 39:1243–1250
https://doi.org/10.1007/s00146-022-01580-2

OPEN FORUM

Social robots as partners?

Paul Healy[1]

Received: 10 October 2021 / Accepted: 6 October 2022 / Published online: 15 October 2022

Abstract
Although social robots are achieving increasing prominence as companions and carers, their status as partners in an interactive relationship with humans remains unclear. The present paper explores this issue, first, by considering why social robots cannot truly qualify as "Thous", that is, as surrogate human partners, as they are often assumed to be, and then by briefly considering why it will not do to construe them as mere machines, slaves, or pets, as others have contended. Having concluded that none of these familiar designations does justice to social robots' still evolving and yet-to-be-defined status, I go on to consider whether engaging in a "relational turn" which prioritises the relationship over the entities in relation, can provide a more satisfactory alternative. In defending this stance, Damiano and Dumouchel (HUMANA.MENTE J Philos Stud 13:181–206, 2020) contend that in addition to foregrounding the possibility that social robots constitute new types of artificial companions which we can find "companionable in different ways", the relational turn has the added advantage of providing a more expansive and productive ethical framework for future research and development in this domain. But on balance, it is far from clear that this approach can circumvent the so-called "dummy-human" problem, the contention that social robots are little more than sophisticated toys with the potential to deceive their users about the prospects for genuine partnership and bonding. Accordingly, I conclude that while Damiano and Dumouchel's shift in "target phenomenology" brings to the fore facets of the interaction that may be difficult to infer by focussing on the relata alone, it is ultimately too "thin" to qualify as a template for robust human-robotic interaction of the sort that could qualify as a genuine partnership. Hence, while their relational turn has its merits in alerting us to the novel possibilities and challenges afforded by this ever-evolving technology and to the limitations of existing frameworks, it is contended that in itself it cannot provide an adequate template for conceptualising what a genuine human–robot partnership might entail. Instead, we need a more expansive approach that can do greater justice to the complexities of human-robotic interaction in its continuities as well as differences from more familiar human (and animal) prototypes.

Keywords Social robots · Thous · Quasi-human partners · Dummy humans · Relational approach

1 Introduction

Although social robots are achieving increasing prominence as companions and carers, their status as partners in an interactive relationship with humans remains unclear. The present paper explores this issue, first, by considering why social robots cannot truly qualify as "Thous", that is, as surrogate human partners, as they are often assumed to be, and then by briefly considering why it will not do to construe them as mere machines, slaves, or pets, as others have contended. Having concluded that none of these familiar designations does justice to social robots' still evolving and yet-to-be-defined status, I go on to consider whether engaging in a "relational turn" which prioritises the relationship over the entities in relation, can provide a more satisfactory alternative. In defending this stance, Damiano and Dumouchel (2020) contend that in addition to foregrounding the possibility that social robots constitute new types of artificial companions which we can find "companionable in different ways", the relational turn has the added advantage of providing a more expansive and productive ethical framework for future research and development in this domain. But on balance, it is far from clear that this approach can circumvent the so-called "dummy-human" problem, the contention that social robots are little more than sophisticated toys with the potential to deceive their users about the prospects for

✉ Paul Healy
phealy@swin.edu.au

[1] Swinburne University of Technology, Melbourne, Australia

genuine partnership and bonding. Accordingly, I conclude that while Damiano and Dumouchel's shift in "target phenomenology" brings to the fore facets of the interaction that may be difficult to infer by focussing on the *relata* alone, it is ultimately too "thin" to qualify as a template for robust human-robotic interaction of the sort that could qualify as a genuine partnership. Hence, while their relational turn has its merits in alerting us to the novel possibilities and challenges afforded by this ever-evolving technology and to the limitations of existing frameworks, it is contended that in itself it cannot provide an adequate template for conceptualising what a genuine human–robot partnership might entail. Instead, we need a more expansive approach that can do greater justice to the complexities of human-robotic interaction in its continuities as well as differences from more familiar human (and animal) prototypes.

2 What kind of partners are social robots?

In the twenty-first century, we are seeing a rapid advance in social robots which "are designed to engage people in an interpersonal manner, often as partners, to achieve social or emotional goals" (Breazeal et al. 2008, 1349). Although social robots are still an emerging technology, this tendency "is gaining currency as more and more interactive personalities appear on the market, as they become steadily more lifelike, more intelligent, more emotional, and more emotionally appealing" (Levy 2010, 93). Moreover, while there is no clear-cut definition of what constitutes a social robot, it is not uncommon to characterise them in quasi-human terms. Thus, for Breazeal, "a sociable robot is able to communicate and interact with us, understand, and even relate to us, in a personal way. It is a robot that is socially intelligent in a human-like way" (Breazeal 2002, in Scheutz 2011, 205). Nonetheless, it remains far from clear what kind of partners they are and what our relationship with them might be. Are they mere things or quasi-persons, or something very different? (Hildt 2019, 2).

2.1 *Social robots as Thous?*

Given that social robots are steadily becoming "more lifelike, more intelligent, more emotional, and more emotionally appealing", and are often so portrayed in popular movies and literature, it seems appropriate to begin by considering the possibility that they could function as quasi-human "Thous", and hence as interactive partners in a genuinely reciprocal I-Thou relationship (cf. Buber 2004/1937; Peltu and Wilks 2010, 279–281). Notably, however, this proposal has met with significant scepticism in some quarters. Thus, in reflecting on the applicability of Buber's I-Thou template to human-robotic partnerships, Turkle's rhetorical question—"The I [is] the person in the relationship, but how could the robot be the Thou?" (Turkle 2010, 6)—casts real doubt on the possibility that a social robot could function as a surrogate human person in a genuinely interactive relationship. Accordingly, let us consider some reasons why, despite popular portrayals, social robots might not qualify as persons and hence as quasi-human partners. While the concept of personhood is itself a complex and contested philosophical phenomenon, it is generally acknowledged that it at least presupposes consciousness, self-awareness, and autonomous agency (cf. Hildt 2019, 2). Can social robots really be said to embody these attributes? In response, it is contended below that not only is this not the case at present but that it may not be achievable, or perhaps even desirable, even in the longer term.

While consciousness is widely considered to be a primary prerequisite for personhood, there is widespread consensus that, however defined, robotic AI does not currently have consciousness, and it remains a contested issue as to whether it ever could, with proponents of "weak AI" more sanguine about this prospect (Hildt 2019; Koch 2019). Notwithstanding ongoing advances in AI which frequently exceed what was previously considered possible, there remain significant obstacles to robots acquiring quasi-human consciousness, not least the fact that scientists and philosophers do not yet understand the "hard problem of consciousness" (Chalmers 1995) in humans well enough to be able to reproduce it in robots. Moreover, the undeniable advances in AI are not reliable indicators of advances in robotic consciousness as there are significant differences between (computational) intelligence and consciousness, most notable perhaps the self-awareness and contextual understanding characteristic of the latter. So, while AI already exceeds human computational power in many respects, this is clearly not tantamount to replicating conscious thought processes. In any case, given differences in the embodiment, circuitry, and architecture between human and artificial consciousness, it seems inevitable that the latter will differ from the former in significant respects.

With regard to the social robot's lack of self-awareness, construed as the ability "to experience a unified first-person self-consciousness", Veruggio and Abney foreground the robot's lack of a Kantian transcendental unity of apperception, noting that there is more to personhood and agency "than mere behavior—there is an interiority, a self who knows what it is like to be someone, in a subjective sense still unexplained by science" (2011, 354–55). The absence of robotic self-awareness is even more apparent if the concept is broadened to include the capacity "to understand their environment, to be cognizant of what they do, to take appropriate and timely initiatives, to learn from their own experience and to show that they know that they have learned and how" (Chatila et al. 2018, cited in Hildt 2019,

1). Moreover, as Veruggio and Abney go on to contend, biological embodiment may also be a requirement for quasi-human personhood. For from an embodied cognition perspective, "the type of body will determine the type of mind that inhabits it." On this view, then, "robots, equipped with mechanical bodies, sensors, and actuators, as well as computational abilities, would have minds, but not human minds—because they lack human bodies" (2011, 356). While weak AI theorists are more optimistic about creating "artificial consciousness" through, for example, computer simulation of brain processes, lacking causal efficacy, these simulations can hardly qualify as quasi-human. Accordingly, they court the possibility that robots "will remain only supersophisticated machinery, ghostlike empty shells, devoid of what we value most: the feeling of life itself" (Koch 2019, 49). For these and related reasons, it seems unlikely that social robots could acquire the consciousness of a substantively humankind. Instead, as in the case of "machine intelligence", progress in this domain is likely to eventuate in artificial consciousness of a distinctive, yet-to-be-defined kind commensurate with its robotic architecture and circuitry (cf. Lyons 2018, 28f.). This conclusion is reinforced by Veruggio and Abney who note that in addition to "our biophysical powers and properties", "the human relationships that we develop and mature from birth until death" also play an integral role in the development of human personhood, such that while "robots may gain capacities that make them our equals or betters in certain ways, … they can never be *Homo sapiens*" (2011, 353). Hence, while it remains difficult to predict the scope of future developments, to the extent that consciousness is a factor, the case for regarding social robots as surrogate human partners, as genuine Thous, is not strong. Their limited autonomy and capacity for independent action compounds the problem. While autonomous agency is also widely regarded as integral to personhood, social robots remain limited in this regard too and, notwithstanding anticipated advances in this domain, the prognosis for attaining quasi-human autonomy in the foreseeable future is not good although, as we shall see, social robot's *perceived* autonomy can indeed lead to attributions of personhood.

Preprograming notwithstanding, even simple robots can be said to have at least some degree of autonomy (Veruggio and Abney 2011, 349), in that they have the capacity for self-movement and can make changes in response to environmental conditions without direct human intervention. Thus, even robotic vacuum cleaners have self-mobility and can change direction to avoid obstacles. Likewise, notwithstanding their circumscribed behavioural repertoires, social robots can make (limited) decisions about what behaviours to exhibit based on "their perceptions of the environment and their internal states" (Scheutz 2011, 206–207). But clearly, this is a far cry from autonomy in the full-blown human sense of a capacity for self-direction and independent decision-making of a sort that would engender ethical accountability. This "much stronger and richer sense of autonomy" hinges on an "agent's active use of its capabilities to pursue its goals", and in the foreseeable future it is unlikely that social robots can acquire "the reflective self-awareness that is prerequisite for accountable, self-governing behavior" (Scheutz 2011, 207). Hence, it is not clear that we can expect social robots ''to be able to do more than produce shallow responses on the basis of a shallow understanding of what they perceive" (Sloman 2010, 194). On this score too, then, social robots cannot currently qualify as Thous, that is, as surrogate human partners in a genuinely interactive relationship, and it is not clear when, or if, these limitations can be overcome. Nonetheless, even the limited autonomy of which social robots are currently capable can impact the dynamics of their interaction with humans in potentially significant ways due to their *perceived* autonomy, which prompts humans "to ascribe intentions to social robots so as "to make sense of their behaviors" (Scheutz 2011, p. 207). We will need to give further consideration below to the ethical problems deriving from this phenomenon in connection with the allegation that social robots constitute a "cheating technology".

A further reason for doubting that social robots could attain quasi-human personhood is provided by continental philosophers like Hans-Georg Gadamer, who famously contends that we are "inescapably more being than consciousness" (1976, 38). In this, he is following his mentor Martin Heidegger who, in his major work *Being and Time*, systematically defends the view that as human beings we have a distinctive ontological constitution that decisively distinguishes us from material objects of all kinds. Heidegger's analysis of the ontological features constitutive of what it is to be a person (termed "existentialia") thus provides an additional touchstone for establishing our distinctiveness vis-a-vis humanoid robots. While it is beyond the scope of this paper to elaborate this dimension, it may be noted that the factors on which Heidegger places special emphasis include the ways in which the meaning and purpose of our very existence are always at issue for us and in which our lives are inextricably bound up with those of others, and in which we confer meaning and significance on things and events in the context of cultural traditions and on the basis of lived experience. Likewise, Heidegger is emphatic about the ways in which lived time "temporalises" itself in a distinctively human way (Heidegger 1962; cf. Gorner 2007, Large 2008). Moreover, since in this view robots do not have, and in principle cannot have, this kind of ontological constitution, our references to robotic intelligence, memory, self-reflective capacity, autonomy and so forth are purely analogical given that, to the extent that robotic AI embodies (or can embody) these attributes, it necessarily does so in a distinctively different way (see, e.g., Ruin 2015a, b on memory).

In sum, then, it would seem that: however, humanoid they may appear, social robots are not, and cannot become surrogate human Thous. On the contrary, as Veruggio and Abney point out, "far from the received biological nature we humans have, a robotic nature will be a choice its engineers make for it." (2011, 357). Hence, as elaborated below, some significant ethical challenges result from their derivative status as so-called "dummy humans" and the consequent potential for the formation of deceptive, unidirectional relationships. This might cause us to reassess whether becoming a person should still be regarded as the desideratum for social robots, notwithstanding the appeal that humanoid robots capable of mirroring human emotions and expressions have for potential users (Bhaumik 2018, 201–202). But if social robots are not Thous, perhaps they occupy the other end of the spectrum as mere machines or slaves?

2.2 *Social robots as machines, slaves, or pets?*

The case for regarding social robots as mere machines or slaves is forcefully put by Bryson, who contends not only that "robots should not be described as persons", but also that they "are fully owned by us". In humanising them, we simply "dehumanise real people" (2010, 63). As a corrective, she maintains, we need to recognise that "robots are tools" which we can and should use "to extend our abilities and increase our efficiency" (73). Hence, it is wrong to attribute autonomy and the related moral responsibility to robots; to do so "is not only false but an abrogation of our own responsibility". Instead, we should be clear that "robots are wholly owned and designed by us", and hence "should be viewed as tools we use to extend our own abilities and accelerate progress on our own goals" (72, 74). But the case regarding social robots as mere machines or slaves is hardly compelling. Thus, firstly, viewing them as mere machines does not seem to do justice to their status as endowed (at a minimum) with (varying degrees of) artificial intelligence and autonomy, nor to their capacity for forming interactive relationships with us. Clearly, this is especially the case for social robots. Likewise, although the term "robot" evidently "originates from the Czech word 'robota', meaning 'slave' or 'serf labour'" (Lyons 2018, 6), there are good grounds for challenging the aptness of the former designation. In particular, while the term "slave" is usually taken to mean a being endowed with agency, autonomy, and personhood which has been subjugated by another, this is not true of social robots which, as we have seen, cannot lay claim to being endowed with the foregoing attributes in any quasi-human sense. Hence, as Veruggio and Abney put it, "in reality, our robots are not (for now, anyway) our 'slaves' in any robust sense, as they have no will of their own" (2011, 352). Moreover, given likely future advances in robotic autonomy and the corresponding defence by some theorists of robot rights, treating them as slaves would arguably "be highly unethical, like racism" (Sloman 2010, p. 185, n. 2).

But if not mere machines or slaves, perhaps social robots are best regarded as (virtual) pets? Thus, as Levy puts it: "For many of those who value their relationship with their pet animal more highly than their relationships with other humans, it would not be very surprising if a virtual pet or a robot were to be regarded in the same vein, supplanting other humans as the most natural objects of affection" (2010, 90). Whether or not virtual pets come to *supplant* humans as primary objects of affection, it would seem that virtual pets can elicit a significant degree of affection and satisfaction as interactive partners. So perhaps, this interaction best mirrors our current relationship with social robots, since it accommodates our affection for them and our readiness to form a relationship with them, while not expecting the same level or degree of autonomy, reciprocity, and ethical responsibility from them as from human partners? Since it is not possible to do justice to this proposal here, suffice it to note in response that social robots are still AI artefacts not biological animals endowed with a significant degree of autonomy, mobility and self-direction. Moreover, the pet analogy also falls short in situations where the social robot elicits additional expectations not usually expected of pets, such as when it functions in the role of carer.

2.3 *Social robots as new, as-yet-undefined kinds of partners*

As contended above, social robots are not Thous, i.e. surrogate human partners, nor, despite ongoing technological developments, are they likely to become Thous. As Veruggio and Abney succinctly put it: although a robot is "a machine that is so complex that it often becomes unintelligible, even to its designer, [it is] always an artificial product of technology, ontologically and irreparably different from a human being" (2011, 362). Neither are they pets, slaves, nor mere machines–though they may rightly be deemed to have affinities with each of these. Instead, the indications are that social robots constitute new types of partners of a still evolving and yet-to-be defined kind which, while having affinities with such familiar types, differ significantly from each of them (in different ways depending on the level of complexity and functioning). As such they present new challenges as well as new opportunities (cf. Taylor et al. 2010, 119–20). Accordingly, it is no longer clear that, or why, emulating a person should be the desideratum for social robots or why approximating a Thou should be the desideratum for partnership. On the contrary, it would seem that social robots serve their purpose better and may be more desirable, even as social companions, if they continue to evolve their own

distinctive, yet-to-be determined identities. Hence, their very (ontological) distinctiveness may prove to be their strength in affording new kinds of companionship, or indeed partnership, while motivating "the possibility of rethinking what companionship might be" (Taylor et al., 114, *passim*).

3 Relational turn: ontological and ethical implications

Recognising the limitations of existing categories, Damiano and Dumouchel (2020) pursue the suggestion that we need to be open to finding social robots "companionable in different ways" by focussing sharply on the relationship rather than the *relata*, the specific entities in relation. In defending this proposal, Damiano and Dumouchel not only embrace the contention that social robots constitute *new kinds* of social partners rather than deficient versions of familiar categories but insist that it is in their interaction that the novel possibilities come to the fore. From this perspective, then, questions about the ontological and ethical status of the *relata*—here, the human and robot, respectively—are eclipsed by a focus on the nature and dynamics of the relationship as such, and in particular the potential it affords for behavioural coordination between robot and human. Deeming it of special significance for the success of interactive exchanges involving social robots, they focus in particular on the emotive dimension, emphasising that their novel "affective loop approach" concentrates "on the robots' capacity to engage humans into affective exchanges" (Damiano and Dumouchel 2020, 190) rather than on its ability to generate emotions or their emotional expressivity as such. Thus, as Balestrieri and Liberati put it, for Damiano and Dumouchel, "The right question is not 'Can robots have emotions?' but 'Can robots establish effective coordination with humans?"' (2020, VI). From this perspective, then, what matters is the effectiveness of the interaction, especially in terms of affective coordination, not the social robot's ontological status or its internal states: "The target phenomenology is different. [The focus] is not individual internal events, but interactive behaviors, the central mechanism [is] affective coordination" (Damiano and Dumouchel 2020, 197). In short, the focus of this approach is on the manner in which and extent to which social robots can effectively coordinate with humans in behavioural terms. In this approach, then, social robots are deemed effective social partners through appropriate behavioural coordination along designated dimensions, irrespective of whether they are regarded as quasi-human Thous, pets, slaves, or mere machines. But while this relational approach may be deemed to have the merit of counteracting an undue preoccupation with the nature of the *relata* by shifting attention to the (dynamics and) effectiveness of the human–robot interaction, it also has its limitations, both ontological and ethical.

Overall, a notable strength is that its shift in "target phenomenology" brings to the fore facets of the interaction that may be difficult to infer by focussing on the *relata* alone. Thus, in particular, it foregrounds the possibility that the satisfaction we derive from affective relationships need not depend on their involving (biological) humans or pets but can derive simply from successful (behavioural) coordination with the entities in question. In so doing, it serves to facilitate recognition that the satisfaction deriving from such relationships is not dependent on the entities in relation being perceived as quasi-human Thous or pets. Rather, these diverse affective relationships can each be worthwhile and satisfying in its own distinctive way. Accordingly, it reinforces the contention that the satisfaction derived from interacting with even highly developed humanoid robots need not depend on their being (mis) taken for quasi-human Thous but may be of a unique sort deriving precisely from their robotic embodiment. From this perspective, then, their "artificial" status could be deemed a strength. On the other hand, a one-sided focus on the relationship at the expense of the *relata* threatens to obscure important information about the status and texture of the relationship, which may have significant implications not just philosophically but also phenomenologically in terms of the human partner's experience and ability to make sense of the relationship. Since perceptions regarding the status of the *relata* will condition our expectations about the texture, dynamics, possibilities, and limitations of the relationship, we need to know something about the ontological status of the *relata* to be able to make sense of the relationship and assess its implications. For, as we shall now consider, failure to take due account of the ontological status of the *relata* can give rise to significant ethical challenges, notably the dummy-human problem. In addition to foregrounding the possibility that social robots constitute new types of artificial companions which we can find "companionable in different ways", Damiano and Dumouchel (2020) contend that their relational turn has the added advantage of providing a more expansive and productive ethical framework for future research and development in this domain. In so doing, they contend that their relational approach can resolve a pressing ongoing ethical problem in this domain, namely, the potential formation of deceptive, unidirectional relationships with so-called dummy humans. We will resume appraisal of their stance following a brief elaboration of the dummy-human problem.

As used by Sloman, the term "dummy-humans" designates: "Products that are intended to be regarded by a user as being like another caring, feeling individual with whom a long-term relationship can develop—even if that is based on

an illusion because the machine is capable only of shallow manifestations of humanity: e.g. learned behaviour rules, such as nodding, smiling, head turning, gazing at faces, making cute noises, or in some cases apparently asking to be comforted, etc." (Sloman 2010, 182–83). As elucidated by Scheutz, there is a real danger that social robots could qualify as dummy humans and hence as "cheating technology" because they induce the formation of unidirectional dependent relationships with social robots which lack the capacity to reciprocate (2011, 211). This danger arises, primarily, because of social robots' *perceived* autonomy which, although actually (very) limited, "prompts humans to ascribe intentions to social robots in order to be able to make sense of their behaviors (e.g., the robot did not clean in the corner because it thought it could not get there)" (Scheutz 2011, 207). It is exacerbated by the endorsement of purely functional conceptions of emotion that "treat emotions as little more than dispositions to produce shallow manifestations that can be recognised as indicating sadness, surprise, being amused, anger, fear, relief, etc." (Sloman 2010, 183). The potential deception arises because, despite appearances, social robots actually have little "to contribute on their end to any relationship"; in effect, they are inept and unable "to partake as a genuine partner". Indeed, they have no notion of "other"; "there are no built-in algorithms for detecting and recognizing people" (Scheutz 2011, 214). Furthermore, when it comes to caring, current robots "do not have the architectural and computational mechanisms that would allow them to care, largely because we do not even know what it takes, computationally, for a system to care about anything" (215). Nonetheless, the evidence is that humans continue to "anthropomorphize robots, project their own mentality onto them, and form what seem like deep emotional yet unidirectional relationships with them" (211), with the potential to "create psychological dependencies that can have serious consequences", while further developments in robotic technology "might prompt people to be even more trusting in them or develop attitudes toward robots that could and likely would be exploited" (216). Hence, the "huge potential for robots to do a lot of good for humans (from elder care to applications in therapy)" must be counterbalanced by considering the "potentially detrimental consequences" of promoting interaction with them (217). As it concerns Damiano and Dumouchel's relational turn, the problem is that focussing one-sidedly on the relationship to the exclusion of the *relata* and thereby obscuring the latter's ontological and ethical status, it could significantly increase the possibility for deceptive human-robotic relationships with potentially damaging consequences, especially in care settings and among other vulnerable populations. But in acknowledging the dummy-human problem, Damiano and Dumouchel characterise the relational turn's ability to respond to it as a core strength. There are two main facets to their response.

Firstly, Damiano and Dumouchel explicitly acknowledge that "robots are objects'', effectively "sophisticated dolls", which while themselves devoid of emotion, nonetheless "induce in users a 'suspension of disbelief', making them appear as humans' peers". Nonetheless, they insist that because "robots can participate in affective coordination dynamics with humans" as outlined above, "robot-human affective interactions are nonetheless authentic" (2020, 194). Hence, in stark contrast to those who view a one-sided focus on affective, or behavioural, coordination as contributing to the dummy-human problem, Damiano and Dumouchel are emphatic that far from constituting a form of "cheating technology", social robots constitute a "new type of affective agents", and hence "a new type of affective partners" participating in "an inter-individual shared affective dynamic" (198–99). In effect, then, from their relational perspective, this new *shared* "affective dynamic" is the (authentic) phenomenon (not to say "reality") demanding our attention, not the fact that the robotic partner is incapable of feeling emotion although the (allegedly deceptive) appearance of so doing that it conveys is integral to sustaining this shared dynamic. Correlatively, Damiano and Dumouchel emphasise that "from an ethical point of view also things now look different, for what is involved is not manipulating the emotions of individual agents, but participating to an inter-individual, shared affective dynamic" (2020, 198). Hence their related contention that their unqualified endorsement of affective coordination as the nub of human-robotic interaction calls for "a radically different ethical position" (199), which they term "synthetic ethics".

In contrast to established ethical stances, "The goal of synthetic ethics is not to judge of what is new … according to a finite set of pre-existing rules, but to discover new rules as needed". It is so named "to emphasize its closeness to 'learning by doing'". So construed, synthetic ethics is characterised by its openness to novelty and experimentation, and to engaging "in critically studying and assessing—case by case, project by project—the effects of human–robot affective coordination" (Damiano and Dumouchel 2020, 201). To this end, synthetic ethics is committed to setting up "experimental scenarios" and observing what happens (when, say, a social robot and a person with "special needs" interact), rather than prejudging the issues using pre-existing presuppositions and standards, ethical or other. In this way, Damiano and Dumouchel maintain that synthetic ethics complements their relational turn not only by deflecting criticisms of social robots deceptively masquerading as dummy humans but also by advancing understanding of how we can create social robots that are "able to stimulate, rather than damage, the cognitive, social and emotional development of their humans partners" (*sic*) (2020, 202). But in

thus insisting on the need for a new ethical approach which transcends traditional utilitarian and deontological thinking, Damiano and Dumouchel omit to explain what it is about a focus on "affective coordination" as such that necessitates this radically different ethical stance or on how specifically it alleviates the potential formation of deceptive unidirectional relationships with dummy humans. Instead, they foreground the deleterious effects that a putative "general condemnation" of social robots as deceptive technology would have on further technological developments in this area, and as an antidote proffer their synthetic ethics, contending that it has the capacity to facilitate, rather than inhibit, research and development in this area (2020, 200–202).

But while such an innovative approach undoubtedly has its strengths, it does not provide a convincing response to the dummy-human problem which, as we have seen, poses a real concern in this context. For, as prefigured above, instead of directly clarifying how a focus on affective coordination might prevent the human partner from being deceived into thinking that the social robot can reciprocate emotionally and otherwise in ways that it demonstrably cannot, Damiano and Dumouchel appear to suggest that their proposed new "target phenomenology" somehow dissolves this problem, thereby rendering it innocuous. But this is not the case because, although the dangers posed by the dummy-human problem can be overstated, the problem cannot be resolved simply by a one-sided focus on the relationship and on affective coordination in particular since, as we have seen, the perceived status of the social robot inevitably gives rise to expectations, including expectations regarding affective bonding, that it may well be incapable of fulfilling. To avoid such misperceptions, it is important to be clear that "there is more to agency than mere behavior" and that while "robots may come to simulate many human abilities", "any simulation always lacks some of the reality of that which it simulates" (Veruggio and Abney 2011, 355). Moreover, Damiano and Dumouchel's contention that acknowledgment of the dummy-human problem implies a "total condemnation" (2020, 200) of this kind of technology as inherently deceptive does not amount to a good reason for sidelining this problem, especially since there is scant evidence of such a generalised condemnation. Hence, while Damiano and Dumouchel perform an important service in foregrounding the need to attend to the relationship and not just the robotic and human entities in relation, their one-sided focus on the former to the seeming exclusion of the latter will not suffice for advancing our understanding of this partnership. Instead, a "both/and" approach which factors in both *relata* and relationship is needed, for without this we cannot arrive at a judicious attribution of ethical responsibilities and assessment of potential dangers. Likewise, while their innovative synthetic ethics clearly has its merits, it will not do to one-sidedly bracket well-established ethical principles and categories in favour of a (purely) experimental approach focussed on "learning by doing". Instead, their proposed synthetic ethics is more justifiably regarded as a complement to, rather than a substitute for, established ethical principles and standards.

In foregrounding the significance of the relationship, Damiano and Dumouchel's analysis reinforces our earlier conclusion that social robots constitute a still evolving and yet-to-be-defined kind of entity which cannot be adequately characterised either as Thous or dummy humans as both designations obscure the social robots' distinctive attributes and capabilities. Damiano and Dumouchel's analysis likewise reinforces the contention that human-robotic interaction poses new possibilities and challenges which transcend the boundaries of existing categories, and which call for a new approach, centred on focused consideration of the nature and dynamics of the relationship. But as we have seen, a one-sided focus on the relationship at the expense of the *relata* will not suffice as the status of the *relata* inherently conditions the expectations and, indeed, possibilities for the relationship. Moreover, important as is the affective dimension, effective human-robotic interaction calls for coordination along multiple dimensions, including the cognitive, physical, and social, so as "to design robots that can successfully play a beneficial role in the daily lives of people" (Breazeal et al. 2008, 1349). Although the I-Thou model with which we started proved untenable as a template, it nonetheless illustrates the need for a more textured, multidimensional, and (mutually) responsive conception of human-robotic interaction. Hence, its other merits notwithstanding, in its one-sided focus on affective coordination, Damiano and Dumouchel's relational turn is too "thin" to serve as a template for robust human-robotic interaction which can "adapt to the changing requirements of their owners over a long period" (Peltu and Wilks 2010, 282). Furthermore, the distinctive possibilities and challenges opened up by Damiano and Dumouchel's relational turn should not cause us to neglect the continuities and lessons from theoretical and empirical research pertaining to the *relata*. Consequently, to consolidate the merits of Damiano and Dumouchel's proposal, we need to adopt an inclusive "both/and" rather than an exclusive "either/or" approach.

4 Conclusion

As contended throughout, the present paper defends the view that social robots constitute still-evolving, yet-to-be defined new technological entities. As we have seen, a strength of this stance is that it acknowledges that social robots do not need to conform to familiar human or animal prototypes to be effective human partners or even to qualify as objects

of our affection (though humanoid features may well be favoured for other reasons). Recognising this should facilitate significant bonding between humans and social robots while allowing the latter to develop their own identity and eschewing criticisms of social robots as deceptive technology. To this end, it is also important to acknowledge and build on the continuities that do exist with familiar entities so as to better conceptualise the possibilities inherent in the emerging social robot technology. Hence, perhaps what matters most is not that we aspire to build social robots that emulate human Thous, but rather that we focus on the design of social robots whose distinctive identity and capabilities can complement and enhance our strengths and thereby our human wellbeing (cf. Breazeal et al. 2008).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

References

Balestrieri M, Liberati N (2020) Living with artificial intelligence. An analysis of moral and philosophical implications of artificial intelligence in our everyday life. HUMANA.MENTE J Philos Stud 13(37):3–4

Bhaumik A (2018) From AI to robotics: mobile, social, and sentient robots. Taylor & Francis Group, Boca Raton

Breazeal C, Takanishi A, Kobayashi T (2008) Social robots that interact with people. In: Siciliano B, Khatib O (eds) Springer handbook of robotics. Springer-Verlag, Berlin Heidelberg, pp 1349–1369

Bryson JJ (2010) Robots should be slaves. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins Publishing Company, Amsterdam, pp 63–74

Buber M (2004/1937) I and Thou. Bloomsbury Academic, London

Chalmers D (1995) Facing up to the problem of consciousness. J Conscious Stud 2(3):200–219

Damiano L, Dumouchel P (2020) Emotions in relation. Epistemological and ethical scaffoldings for mixed human-robot social ecologies. HUMANA.MENTE J Philos Stud 13(37):181–206

Gadamer H-G (1976) Philosophical hermeneutics. University of California Press, Berkeley (**Translated by David Linge**)

Gorner P (2007) Heidegger's being and time: an introduction. Cambridge University Press, Cambridge

Heidegger M (1962) Being and time. Harper & Row, New York (**Translated by J. Macquarrie and E. Robinson**)

Hildt E (2019) Artificial intelligence: does consciousness matter? Front Psychol 10(1535). https://doi.org/10.3389/fpsyg.2019.01535. https://www.frontiersin.org/article/10.3389/fpsyg.2019.01535

Koch C (2019) Will machines ever become conscious? Sci Am 321(6):46–49. https://doi.org/10.1038/scientificamerican1219-46. https://www.scientificamerican.com/article/will-machines-ever-become-conscious/

Large W (2008) Heidegger's being and time: an Edinburgh philosophical guide. Edinburgh University Press, Edinburgh

Levy D (2010) Falling in love with a companion. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins Publishing Company, Amsterdam, pp 89–93

Lyons S (2018) Death and the machine: intersections of mortality and robotics. Springer Singapore, Singapore

Peltu M, Wilks Y (2010) Summary and discussion of the issues. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins Publishing Company, Amsterdam, pp 259–286

Scheutz M (2011) The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey GA (eds) Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, pp 205–221

Sloman A (2010) Requirements for artificial companions: it's harder than you think. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins Publishing Company, Amsterdam, pp 179–200

Taylor A, Jain A, Swan L (2010) New companions. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins Publishing Company, Amsterdam, pp 107–120

Turkle S (2010) In good company? On the threshold of robotic companions. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins Publishing Company, Amsterdam, pp 3–10

Veruggio G, Abney K (2011) Roboethics: the applied ethics for a new science. In: Lin P, Abney K, Bekey GA (eds) Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, pp 347–363