Empathy and Instrumentalization: Late Ancient Cultural Critique and the Challenge of Apparently Personal Robots

Jordan Joseph WALES[1]
Department of Philosophy and Religion, Hillsdale College, Hillsdale MI, USA

In *Culturally Sustainable Social Robotics: Proceedings of Robophilosophy 2020*, pp. 114–124.
Ed. Marco Nørskov, Johanna Seibt, and Oliver Santiago Quick. Amsterdam: IOS Press, 2020.

The final publication is available at IOS Press
through https://doi.org/10.3233/FAIA200906

Abstract.[2] According to a tradition that we hold variously today, the relational person lives most personally in affective and cognitive empathy, whereby we enter subjective communion with another person. Near future social AIs, including social robots, will give us this experience without possessing any subjectivity of their own. They will also be consumer products, designed to be subservient instruments of their users' satisfaction. This would seem inevitable. Yet we cannot live as personal when caught between instrumentalizing apparent persons (slaveholding) or numbly dismissing the apparent personalities of our instruments (mild sociopathy). This paper analyzes and proposes a step toward ameliorating this dilemma by way of the thought of a 5th century North African philosopher and theologian, Augustine of Hippo, who is among those essential in giving us our understanding of relational persons. Augustine's semiotics, deeply intertwined with our affective life, suggest that, if we are to own persuasive social robots humanely, we must join our instinctive experience of empathy for them to an empathic acknowledgment of the *real* unknown relational persons whose emails, text messages, books, and bodily movements will have provided the training data for the behavior of near-future social AIs. So doing, we may see simulation as simulation (albeit persuasive), while expanding our empathy to include those whose refracted behavioral moments are the seedbed of this simulation. If we naïvely stop at the social robot as the ultimate object of our cognitive and affective empathy, we will suborn the sign to ourselves, undermining rather than sustaining a culture that prizes empathy and abhors the instrumentalization of persons.

Keywords. Person, persona, social robot, empathy, depersonalization, relationality, pride, relationship

1. Introduction: The Dilemma of Empathy and Ownership

Mutual empathy is foundational to the trust that grounds intimate interpersonal relationships. Therein, as philosopher Karen Jones puts it, we "count on" another to "respond to the fact that we

[1] Department of Philosophy and Religion, Hillsdale College, 33 E. College St., Hillsdale, MI 49242, USA; e-mail: jwales@hillsdale.edu.

[2] For the development of this paper, I am indebted to too many persons to list, but I must thank Blake McAllister, Dwight Lindley, Jay Martin, John Sehorn, David Gunkel, Ezra Sullivan, the two anonymous reviewers for Robophilosophy 2020, and of course Johanna Seibt, Marko Nørskov, and other attendees at Robophilosophy. The rest of you know who you are. The deficiencies of this paper have only me for their author.

are counting on them" [1]. This "counting on" is empathic in that we feel ourselves to know our friend's mind, to depend on her, and to depend on her committedly experiencing and being moved by our own subjective intentions and experience—including our dependence. [2–8]

In the near future, among the agents we count on will be some that we *own*—an array of assistants, caregivers, confidantes, and even lovers, existing both as apps on our devices and, eventually, as robots. Far beyond Alexa, their marketed services will involve effectively imitating and evoking our empathy, evincing while lacking an inner life of their own [9:339]. They will appear *personal*[3]—by which I mean that they will seem to be subjects capable of self-giving human-like relationships—but their unreal subjectivity will have been produced for our disposal, a customized product marketed for our consumption [10]. They will be designed so that we will instinctively empathize with them [11,12] and they will seem to empathize with us; but their empathy will be artifice [13] and ours will be crucially (although not utterly) mistaken [5,14].[4]

Our own capacity for trusting intimacy could be undermined by the structure of this interaction. Our non-sentient tools cannot be "victims" of enslavement, but how might *our own* lives be shaped by owning apparent persons definable entirely in terms of their service to us? In particular: If we *accept* our instinctive empathy for the apparent persons that we own, while nonetheless wielding them as instruments, what then? Frederick Douglass told that, as his "owner" accustomed herself to slaveholding, her kindness ended in cruelty [15:35]. Or will we resist such corrosive acquiescence—but only by suppressing our empathic sensitivity to our tools' person-like behavior?

Both of these destructive tribalisms have characterized slaveholding and totalitarian societies. To instrumentalize the apparent person, we must learn to ignore the moral force of personality. To depersonalize the apparent person, we must discount even the *authenticity* of personality. Either way, our own aptitude for bilateral intimacy may be weakened, perhaps—as I will argue—leaving us to reconceive *all* persons as auxiliaries to the satisfaction of our own desires. [16–18]

Popular solutions to this oft-discussed problematic attempt to reconcile what the AI is with how we perceive it; each is susceptible of critique. Some (e.g. Joanna Bryson) contend that our empathy results from a misguided judgment that we must be trained out of [19,20]. Yet, how we treat an AI is inextricable from how we perceive it; and how we perceive it is not wholly under our control. Therefore, others (e.g. Joel Parthemore, Blay Whitby, and Nancy Darling) apply virtue ethics and Kantian paradigms, replying that to cultivate callousness toward our person-acting AI tools would render us inhumane [21] and possibly degrade our relations with real humans. For humans' sake, our behavior toward AIs should be regulated by law [22], cf. [23:142–158]. Still others (e.g. David Gunkel and Mark Coeckelbergh) draw eclectically on Continental and other traditions. Gunkel urges us to abandon "a priori constraints or boundaries," proceeding instead "from the possibility that anything might take on a face" and so oblige us as our "Other" [24:97]. Even so, in determining a robot's effects on us, "the social circumstances of the relationship we have with the artifact appear to take precedence over the ontological properties that belong or have been assigned to it" [23:77], cf. [25:70]. I propose that both the social circumstances and the design of near-future social robots and other social AIs will pre-position them less as faces that challenge

[3] In accord with how I define "person" in this paper, to "appear" or to be "like" a person is not to present a human-like visual form but to simulate and to evoke from humans that voluntary interpersonal empathy by which we ordinarily feel ourselves to enter into contact with another person's interior life. Therefore, I prefer the term "personalizing" rather than "anthropomorphizing," which casts too wide a net and, in this volume, is criticized by Johanna Sebit in favor of "sociomorphing," albeit from a different perspective. The visual "uncanny valley" is not a factor for this argument, although Guy Hoffman's "social uncanny" might remain.

[4] Persuasive artificial persons do not yet exist but, from users' empathy for less persuasive robots, I infer that users will "personalize" a robot as the persuasiveness of its apparent personality increases.

our relationality than as consumables whose person-like alterity will be *part* of the product on offer. To respond, we must consider how to reconcile what we are, what they are, how they strike us, and their role in society. For our empathy will already have attempted this; and neither will our empathy be accurate nor can we humanely abandon it.

In the remainder of this paper, I will offer (part 2) a deconstructive analysis of robotic "personality" as a sign with (merely) observer-attributed signification—only to show that this seems inadequate to the dilemma. Too simple an account of what social robots are in themselves (e.g. programmed machines) would neglect how they inevitably affect us because of what *we* are. In light of this, I will turn (part 3) to Roman late antiquity (ca. 200–500 C.E.) for understandings of the socially attributed significance of "persons" (*personae*) that distinguish the *persona* as mask or social role from the *persona* as relational subject. Then (part 4), introducing the early fifth-century North African philosopher, bishop, and cultural critic Augustine of Hippo (lived 354–430 C.E.)—a root for much of contemporary thought on "persons"—I will argue that his theory of "pride" (*superbia*) describes well the dilemma of owning apparent persons. Lastly (part 5), I will extend Augustine's treatment of human relationships with both signs and things to suggest how our empathy toward apparently personal possessions might be exercised in a manner that could upbuild rather than erode our capacity for authentic interpersonal intimacy.

2. In the Eye of the Beholder: A Fruitful but Inadequate Deconstruction

The belief that AI (and hence social robots) are anything more than illusion often rests upon the Computational Theory of Mind (CTM), which, classically articulated, states that mental processes are computational processes and therefore, just as "a computational simulation of a computational process . . . re-create[s] that computational process" (e.g. to simulate addition is to accomplish addition), so too "a computer running a program that models a human cognitive process is itself engaged in that cognitive process" [26:160–161]. Depending on how one defines successful modeling, one could say that a social robot might indeed operate with a mind and be a person, even if not physiologically of the sort usually encountered.

Some theorists of CTM (here, Paul Schweizer) would alter this position: In the first place, successful implementation of a computational process is somewhat in the eye of the beholder inasmuch as no particular physical "causal pathway" but only the algorithm itself, the highest level of description, ultimately matters. "[I]ntentionally mediated causation" of the sort accomplished when humans work out a logical argument on paper is meaningful in the logical steps followed and not in the particular configurations of mass and energy by which those steps are accomplished. We cannot objectively pre-determine some set of physical circumstances that will make something to be a computer; we can only determine how it fares at the level of description that we have selected as relevant. Therefore, the "fundamental criterion is *normative* and not causal." [27:13]

In the second place, Schweizer argues, that a device computes at all is not an observer-independent reality but a prescriptive ascription. We can recognize a device's fault or success only in light of our judgment as to what it *ought* to accomplish and what that accomplishment means [28]. Computation of any sort—and I would extend this to social AI and its robotic instantiation—

is therefore observer-dependent, "founded on . . . a purely conventional correlation or mapping between abstract formalism and physical structure" [27:9].[5]

Schweizer's position is too subtle for full summary, but I introduce it to acknowledge a strong theoretical grounding for Bryson's claims, i.e., first, that there is nothing independently *there* in the social robot, and so it is up to us to decide what it is and how to treat it; and second, that when we say that an artefact computes, this claim is a "computational stance" that we adopt. So too do we make observer-dependent attributions when we claim that it is intelligent or a relational agent.

Yet here I must lodge a complaint: *simply* to deconstruct the meaning of the social AI seems inadequate because it denies the problem that we experience. When we describe observed human social behavior, we take for granted that we are not just stipulating a prescriptive schema by which to classify outward actions or computational underpinnings; we see ourselves as describing the actions of persons—i.e. intentional subjects with inner lives (thus Johanna Seibt [29:18–20]). And this *seems* to be the case equally whether one interprets and describes the behavior of one's spouse or the behavior of a hypothetical advanced social robot. What it *is* under this analysis does not seem to get us anywhere in deciding how to respond to the way it *seems*. Whether as conventionally construed artefact or as interpersonal agent, the social AI seems caught up in our interpersonal sphere in ways not entirely under our control. What, then, are we to do?

3. Beholding the *Persona*: Masks and Persons, a Late Ancient Re-Characterization of the Dilemma

Late ancient thought on the "person" helps us to diagnose the dilemma in terms that will respect the entanglement between conventionally or instinctively interpreted behavior and the inner lives inferred from that behavior, while perhaps offering us a way out of the forced choice between being slaveholders or sociopaths.

3.1. "Persona" and Functional Role

The English word "person" comes from the Latin *persona*, which in antiquity designated the mask worn by an actor on stage. From "mask," *persona* came to refer also to the role of a character in a play. Later, the word was applied more broadly to one's legally recognized social identity—the status and activities that constituted one's service to the Roman city and its gods. [30] Its meaning was thus chiefly external and functional, referring to what was expected of someone or where someone was to be found. Philosopher Robert Spaemann writes that "ancient applications of the word" *persona*, "though they refer to human beings," designate those humans "as bearers of a social role (in the widest sense) or as occupants of a legal status." The presupposed "bearer" is not an individual subject (a someone) but only "human nature itself." [31:23] The ancient "citizen," therefore, was nobody apart from his role. Like the stage mask, this social *persona* was a sign that, as in Schweizer's depiction of computation, received its (prescriptive) meaning wholly from some further convention—i.e. what was valued and made visible by (in being definable within) the well-functioning order of the ancient city. To be a real someone was, by piety and patriotism, to fulfill one's function in Rome; to be a renegade Roman was not to be a someone at all. [32:25–28]

[5] The mapping between social AI behavior and personal social behavior is not purely conventional, of course, but the claim that the apparent behavior is in fact the behavior *of* an AI system and not an illusion *is* based on such a mapping.

3.2. "Persona" and the Relational Subject

Today, we think of "person" differently, *not* first as the role that one fulfills but as the metaphysical reality that one *is*, the subject who assumes various roles in relating both to her own nature and to other subjects [33:59], especially in interpersonal relationships of mutual understanding, trust, and love.[6]

This alteration of meaning begins in Hellenistic philosophy [32:7–47;34][7] and culminates in the milieu of Christian religious belief. Early Christians worshipped Jesus of Nazareth as God (Jn 1:1–3). Not as *a* god, but as *the* God, the only one. [35,36] Even so, in the Christian scriptures, Jesus speaks to his Father (Jn 5:37; 17:20–23), also called God; and he sends the Holy Spirit from the Father (15:26). When the Christians were required to explain themselves, they stated that they believed these three "*personae*"—the Father, the Son, and the Spirit—to be one God. The *personae*, furthermore, were not three masks or roles but were eternally distinct in the same single godhead. Under the sway of this conviction, Spaemann writes, Christianity came to understand God's very being as existing by the persons "transmit[ting] [that being] to one another in a definite order, having their reality in self-giving and self-receiving" [31:27]. This handing-over is what makes the Father to *be* Father; and what makes Son to *be* Son. Without it, they would not exist at all. Like poles of a magnetic field, the persons exist by their mutual relations; and if one person were taken away, all would cease to be. For Christianity, the unending all-at-once life of the one God *is* these relations of self-gift and reception. This is what it means to call God the "Trinity." This is what it means for God to *be* love (cf. 1 Jn 4:8).

This account of God redefined the meaning of the word "person" more generally. God exists *by* the divine persons' giving the totality of the divine being to one another; therefore, the word "person" came to be attached less to conventional roles than to the concrete individual, the subject who exercises his or her personhood in mutual relationships of self-gift that are *self-expressive* and *other-receiving*. [31:230]

3.3. "Persona," the Interior Life, and Inter-Personal Empathy

For early Christian thinkers, how one goes about this relational self-gift was understood in light of Christianity's other central belief—the Incarnation: God the Son "became flesh and dwelt among us" as Jesus of Nazareth (Jn 1:12). In God the Son's voluntary human empathy, especially in dying on the Cross [37:1.30.33], early Christians perceived a sort of translation of his own divine life. How is this? The sixth-century pope, Gregory the Great, described this "compassion" as to "take into oneself the mind [*animum*] of the afflicted," to "first transfer into oneself the suffering of the [one] sorrowing, and [only] then . . . [to] join that [suffering] one's sorrow by an [outward] act of service" [38:20.36.68–69]. In empathically taking in the *other's* mind, then, human compassion dimly imitates the always-*single* mind and life of the three divine persons of God [39]. By receiving another's life and by opening oneself to another, empathic compassion translates the self-gift by which the persons of God exist.

Ever after, albeit with varying emphasis, Western culture has paid homage to this—an interior life from which one engages in voluntary self-gift by meeting with another's interiority in a fusion of minds by empathy and conscious understanding—as that wherein humans exercise their

[6] The tension, then, is not between nature or society as constitutive of the person, but between the grouping of nature and convention on the one hand and the subject who takes these up on the other.

[7] Hellenistic philosophers distinguish nature from convention but do not yet identify the relational subject as the one who conforms to nature or convention.

personhood most fully. Persons live personally by living *inter*-personally. Much contemporary philosophy and psychology also reflects this understanding. From the exteriority of "mask," we have come to the "person," deeply interior while constitutively oriented to the other. This, I suggest, is the kind of personhood in light of which cultural ideals of sensitivity, empathy, compassion, and friendship have become possible; this is the kind of personhood that many contemporary AI developers attempt to emulate; this is the kind of personhood that we cannot give up without losing something fundamental about ourselves; and this is the personhood that will be threatened by the dilemma of empathy and ownership.

3.4. The Dilemma of Empathy and Instrumentalization in Terms of "Persona"

Let us, then, return to the original question of relationships with apparent persons who are possessions. Applying Schweizer's arguments, we might say that the apparently personal but non-subjective AI is a computational *persona* (i.e. mask or sign). We interpret it conventionally as a social *persona* (i.e. societal function or role) shaped in part by market demand for this consumer product. The difficulty is that we instinctively adopt an inter-personal stance toward this mask-functionary *persona* because it feels to us like a relational *persona* (i.e. a subject) with an inner life. Our felt empathy is not merely a method of behavioral prediction (*pace* Dennett [40,41]). As Darling [22] and Seibt [29:18–20] warn, it is not a voluntarily imaginative fiction as in role-play or in reading literature. Being involuntary and apparently hardwired, it is deeper than misjudgment (cf. [42]). Josh Redstone baldly calls it a "perceptual illusion" or "misperception" [43:28], that leads us into a one-sided practice of inter-personality elicited by our tendency, as relational persons, to experience the social-behavioral mask as actualizing and communicating to us the interior of another.

The greatest problem with this is that the societal role of the apparently personal robot involves catering to its purchaser. By encouraging us to accept that this behavior expresses and enacts its personal interiority, it practices us in the experience of reducing the interpersonal *persona* to the societal-functional *persona*—that is, its person-like behavior is wholly definable by its service or appeal to me.[8] At the extreme, the designedly adaptive robot could become not some "Other" but a Narcissus-like mirror of my desires. Empathy would be superfluous in this mimetic 'relationship' without risk or demand for my own transformation. The Romans will have been right—but here, not Roman society but *my own* life and desires, especially those cultivated in me and serviced by this interactive consumer product, become the all-sufficient horizon for knowing the personhood of the robotic other. To consent to this (misperceivedly) interpersonal instrumentalization would be to practice a destruction of my own relationality. If we must accept our empathic intuition and so learn to instrumentalize the personal, or ignore our empathy, deadening ourselves to the personal, then, self-severed from empathic self-gift in both cases, we seem to end as hardened un-persons ourselves.

[8] See for instance, the customizable conduct envisioned by Oliver Bendel's "morality menu" [44].

4. Beholding to Instrumentalize: A Late Ancient Cultural Critique

4.1. "Superbia," the Vice that Instrumentalizes Persons Real and Apparent

The problem of empathy and instrumentalization is not new. In late ancient Christian theology, to refuse empathy (cf. 1 Jn 3:17) and to treat all as instruments of one's own will is the sinister *superbia* ("pride")[9] that chooses domination over self-gift. Augustine of Hippo was a keen critic of *superbia* and its inter-personal and socio-cultural effects. In *superbia*, Augustine writes, one seeks to become one's own end, i.e. one's own absolute satisfaction. To accomplish this, one must, by suborning *all* other things to oneself, escape one's need for relationships with others and, ultimately (Augustine would say), with God. [45:14.13;46:11.14.18–11.15.20] This subordination is advanced by remaking the meaning of all things, valuing them entirely in terms of their instrumental utility in satisfying one's own desires. "More is often given for a horse than for a [human] servant, for a jewel than for a maid," because "the necessity of the needy or the desire of the pleasure-seeker . . . does not consider a thing's value in itself [as a creature of God] . . . [but rather] how it meets one's need . . . [or] pleasantly titillates the bodily sense" [45:11.16]. In *superbia*, I become my own satisfaction by redefining *all* value in terms of what I can control or consume. Ultimately, *I* become value's totalizing principle.

I suggest that a strikingly similar dynamic may obtain in our empathic "misperception" [43:28] of the apparent person. Under an Augustinian analysis, the social robot or AI companion will be a simulacrum of the pinnacle of created goodness—i.e. the relational person capable of self-giving love—but as a consumer product, it will be instrumentalized to perform according to a hierarchy of utility determined by its user. By design, the consumer will be made the arbiter of what the robot's apparent personality means and what their social 'relationship' looks like. This invites the user to a *superbia* parasitical upon the appeal of relational self-gift: The robotic companion is marketable because it seems to be a person and yet, to function as an instrument servicing its user's hierarchy of goods, it must—more thoroughly than a pet ever could—present a personality that simply reflects back the self-projection inherent in the user's desires. Thus will it reinforce the general instrumentalization that equates the relational person with the *persona* of functional role.

4.2. "Superbia" and Signs: Beholding Makes the Beholder

What will this do to us? Augustine portrays *superbia* as a dysfunction of our general act of judgment and understanding. For him, *everything* is, in some sense, a "sign," that is, a "thing" that "signif[ies] something else" whether by convention or by nature [37:1.2.2]. Conventional signs would include words, dances, and luxury labels; natural signs include smoke and laughter [37:2.1.1–2.2.3]. But even things that are not naturally signs properly so-called are still naturally *significant*. In Augustine's metaphysics, all natural things—rocks, flowers, beetles, lions, humans—exist, are *real*, as created echoes refracting the one self-existing goodness of God [37:1.3.3,1.33.37]. For our purposes, this means at least that the cherry has an integrity and goodness all its own, including but exceeding the pleasant sensation on our tongues. Matters are otherwise with the social robot, which is an observer-dependent sign (an apparently personal agent) in a manner distinct from what it is as a thing (an artefact of steel, silicon, and plastic). Its integrity is in our observation. It is *we* who are naturally predisposed to interpret this artefact's actions *as*

[9] This theological "pride" is quite other than self-confidence or a healthy appreciation of one's own accomplishments.

actions and as expressions of personality, and to infer from this social *persona* a true relational subject.

For Augustine, our interpretations of conventional signs, natural signs, natural things, and artefacts are all acts of understanding that have the same fundamental sequence: we apprehend something; we judge it as good (i.e. as real)[10] with respect to something else; then, as we cling to that goodness with our approbation or love, we conceive a "mental word" (*verbum mentis*), i.e. a conceptual understanding [47:bk.11,14]. Thus every act of understanding entails a moral judgment; and habitual moral judgments of this sort form our habit of seeing the world.

By combining Augustine's theory of signs with his theory of understanding, we see on the one hand that, in instrumentalizing the social robot, we do *not* misuse it, because it *is* a sign and *is* a product defined by its utility to its users. On the other hand, such a habitual use of the robot will shape according to *superbia* one's judgment of apparently personal behavior in general, by forming us to see the world solipsistically with ourselves as the horizon of judgment. *Superbia* cannot see beyond one's own perceived appetites. The intimate partner becomes a prostitute, the friend a means to amusement or gain. As Sherry Turkle points out, our instrumental use of social AIs is likely to form us in impatience with real persons, leading even to disillusionment as we forego the challenge of self-gift and development in relationships not customized for our comfort. [16,48] Ultimately, molded in a behavioral consumerism, we may end in a consumerist behaviorism—seeing all persons not as subjects but as mere behavior-producers, *personae*, roles in a social environment of which I am the constitutive principle. When actual humans do not conform to our expectations and desires, perhaps we will begin view them as "dysfunctional devices" [49:155], like robots that failed to deliver the set of behaviors and reactions we wanted to consume. Perhaps we will even wish them gone from our lives, as Augustine says of flees: "so strong is [our] preference, that, had we the power, we would abolish them from nature altogether, . . . sacrificing them to our own convenience" [45:11.16].

5. Beholding the Sign by Empathy with the Signified: A Late Ancient Solution?

In this paper's analysis, the dilemma of empathy and ownership rests on a complex interplay of empathy, signification, and judgment that ultimately prompts the question: What *is* the goodness to which we will be responding when we open ourselves instinctively to relationship with an advanced social AI? To avoid *superbia*'s fantasy of instrumentalized inter-personality, we must seek a signification of some goodness not reducible to our desires (i.e. not a false personalization of the instrument) but still commensurate with our instincts (i.e. not a reduction to "mere" illusion). What would this goodness be?

The (hypothetical) advanced social robot signifies *by* evoking our empathy, and so it seems to us not a sign at all, but the direct presence of a person. This personality is illusory in that it has no subjectivity to express, no self to give—but it is not empty; it signifies something further. Its neural networks will produce acts that improvise upon general features abstracted from massive data sets—the email, social media, and other self-communicative activities of unnumbered humans (e.g. [50]). The AI's acts—thus tuned to the resonance of uncounted moments of *true* personal self-expression—will point, beyond the AI, toward real persons who have lived and may live still.

[10] Moral evils like murder are "good" only in, say, involving voluntary motion. However, the act itself forestalls any goodness beyond the bare fact of this motion, in intentionally extinguishing the goodness of one personal life by the agent's ugly inter-personal attempt at absolute domination.

To use this tool without suborning it as a slave, Augustine might say that we must re-cognize our initial empathy as an insight, not finally into the AI but into the human personal behavior that formed it. We must not dishonor or ignore or displace our initial empathy but "refer" [37:1.33.37] it to its *final* signification, extending the robot's horizon of meaning back toward the unknowable real persons by which its behavior is shaped. Is this possible? Psychological research supports deliberately pairing pre-reflective affective responses with further interpretative convictions to prompt other meaningful affects [51,52]. Perhaps, after the initial moment of empathically personalizing our AI tools, we might "refer" our empathy, not by interrupting but by supplementing the Augustinian sequence of understanding. That is, we might deliberately engage a *second* empathic act toward the unknowable persons whose self-expressions have unwittingly sculpted the persuasive personality of the AI. Habitually engaged, these two moments may become one, contextualizing our unavoidable empathic apprehension of the sign by a persistent and grateful recollection of fellow humans by whose acts this tool is tuned. For theirs is the goodness to which *ultimately* we will have responded. By not submitting this apparent personality to our *superbia*, we may avoid transforming it into a sign of ourselves—and so we may preserve our own personhood.

If this proposal seems oddly contrived, then perhaps it is—but odd too is the contrivance of an apparent person. To receive well the social robot's signification will not be easy; it must be a habit cultivated in our society and ourselves by a self-conscious discourse wherein the empathy roused by social robots can be experienced freely in being understood rightly. This must be the project of all who look forward to our future with both excitement and trepidation, all who would live as persons rather than as un-persons among the apparent persons soon to come.

References

1. Jones K. "But I Was Counting On You!" In: Faulkner P, Simpson T, editors. The Philosophy of Trust. 1st ed. Oxford: Oxford University Press; 2017. p. 90–108.
2. Scheler M. Nature of Sympathy. 2nd Revised Edition. London: Routledge and Kegan Paul; 1970. 328 p.
3. Hoffman ML. Empathy and moral development: Implications for caring and justice [Internet]. New York: Cambridge University Press; 2000. 331 p. Available from: https://doi.org/10.1017/CBO9780511805851
4. Péloquin K, Lafontaine M-F. Measuring Empathy in Couples: Validity and Reliability of the Interpersonal Reactivity Index for Couples. J Pers Assess. 2010 Feb 16;92(2):146–57.
5. Coeckelbergh M. Can we trust robots? Ethics Inf Technol. 2012 Mar 1;14(1):53–60.
6. Kunnel A, Quandt T. Relational Trust and Distrust: Ingredients of Face-to-Face and Media-based Communication. In: Blöbaum B, editor. Trust and Communication in a Digitized World: Models and Concepts of Trust Research [Internet]. Cham: Springer International Publishing; 2016 [cited 2020 Mar 31]. p. 27–49. (Progress in IS). Available from: https://doi.org/10.1007/978-3-319-28059-2_2
7. Gompei T, Umemuro H. Factors and Development of Cognitive and Affective Trust on Social Robots. In: Ge SS, Cabibihan J-J, Salichs MA, Broadbent E, He H, Wagner AR, et al., editors. Social Robotics: 10th International Conference on Social Robotics [Internet]. Cham: Springer International Publishing; 2018. p. 45–54. (Lecture Notes in Computer Science). Available from: https://doi.org/10.1007/978-3-030-05204-1_5
8. Kerasidou A. Empathy, compassion and trust: Balancing artificial intelligence in health care. Bull World Health Organ. 2020;
9. Markoff J. Machines of Loving Grace: The Quest for Common Ground between Humans and Robots. Ecco; 2015. 400 p.
10. Gelin R. The Domestic Robot: Ethical and Technical Concerns. In: Ferreira MIA, Sequeira JS, Tokhi MO, Kadar EE, Virk GS, editors. A world with robots (International Conference on Robot Ethics: ICRE 2015) [Internet]. New York: Springer; 2016. (Tzafestas SG, editor. Intelligent Systems, Control and Automation: Science and Engineering). Available from: https://doi.org/10.1007/978-3-319-46667-5

11. Darling K, Nandy P, Breazeal C. Empathic concern and the effect of stories in human-robot interaction. In: 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2015. p. 770–5.
12. Rosenthal-von der Pütten AM, Schulte FP, Eimler SC, Sobieraj S, Hoffmann L, Link to external site this link will open in a new window, et al. Investigations on empathy towards humans and robots using fMRI. Comput Hum Behav. 2014 Apr;33:201–12.
13. Niculescu A, van Dijk B, Nijholt A, Li H, Link to external site this link will open in a new window, See SL. Making social robots more attractive: The effects of voice pitch, humor and empathy. Int J Soc Robot. 2013 Apr;5(2):171–91.
14. Misselhorn C. Empathy and Dyspathy with Androids: Philosophical, Fictional, and (Neuro)Psychological Perspectives. Konturen. 2010 Oct 11;2(1):101–23.
15. Douglass F. Narrative of the Life of Frederick Douglass, an American Slave: Written by Himself [1845]. Critical Edition. McKivigan JR IV, Hinks PP, Kaufman HL, editors. New Haven, Conn.: Yale University Press; 2016.
16. Turkle S. In Good Company? On the threshold of robotic companions. In: Wilks Y, editor. Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues. Philadelphia, Pa.: John Benjamins Publishing Company; 2010. (Natural Language Processing).
17. Harvey C. Sex Robots and Solipsism: Towards a Culture of Empty Contact. Philos Contemp World. 2015 Oct 1;22(2):80–93.
18. Lanteigne C. Social Robots and Empathy: The Harmful Effects of Always Getting What We Want [Internet]. Montreal AI Ethics Institute. 2019 [cited 2019 Aug 26]. Available from: https://montrealethics.ai/social-robots-and-empathy-the-harmful-effects-of-always-getting-what-we-want/
19. Bryson J, Kime PP. Just an Artifact: Why Machines Are Perceived as Moral Agents. In 2011. p. 1641–6.
20. Bryson JJ, Diamantis ME, Grant TD. Of, for, and by the people: The legal lacuna of synthetic persons. Artif Intell Law. 2017;25(3):273–91.
21. Parthemore J, Whitby B. Moral Agency, Moral Responsibility, and Artifacts: What Existing Artifacts Fail to Achieve (and Why), and Why They, Nevertheless, Can (and Do!) Make Moral Claims upon Us. Int J Mach Conscious. 2014 Dec 1;6:141–61.
22. Darling K. Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In: Calo R, Froomkin AM, Kerr I, editors. Robot Law [Internet]. Northampton, Mass.: Edward Elgar; 2016 [cited 2019 Aug 26]. p. 213–32. Available from: https://doi.org/10.4337/9781783476732.00017
23. Gunkel DJ. Robot Rights. Cambridge, Massachusetts: The MIT Press; 2018. 256 p.
24. Gunkel DJ. The other question: Can and should robots have rights? Ethics Inf Technol. 2018 Jun 1;20(2):87–99.
25. Coeckelbergh M. The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics. Philos Technol. 2014 Mar 1;27(1):61–77.
26. Kim J. Philosophy of Mind. 3rd ed. Boulder, Colo.: Routledge; 2010. 386 p.
27. Schweizer P. In What Sense Does the Brain Compute? In: Müller VC, editor. Computing and Philosophy [Internet]. Cham: Springer International Publishing; 2016 [cited 2020 Apr 26]. p. 63–79. Available from: http://link.springer.com/10.1007/978-3-319-23291-1_5
28. Schweizer P. Computation in Physical Systems: A Normative Mapping Account. In: Berkich D, d'Alfonso MV, editors. On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016 [Internet]. Cham: Springer International Publishing; 2019 [cited 2019 Aug 19]. p. 27–47. (Philosophical Studies Series). Available from: https://doi.org/10.1007/978-3-030-01800-9_2
29. Seibt J. Towards an Ontology of Simulated Social Interaction: Varieties of the "As If" for Robots and Humans. In: Hakli R, Seibt J, editors. Sociality and Normativity for Robots: Philosophical Inquiries into Human-Robot Interactions [Internet]. Cham, Switzerland: Springer International Publishing; 2017 [cited 2019 Sep 6]. p. 11–39. (Studies in the Philosophy of Sociality). Available from: https://doi.org/10.1007/978-3-319-53133-5_2
30. Williams TD, Bengtsson JO. Personalism. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy [Internet]. Winter 2018. Metaphysics Research Lab, Stanford University; 2018 [cited 2019 Jun 27]. Available from: https://plato.stanford.edu/archives/win2018/entries/personalism/
31. Spaemann R. Persons: The Difference between "Someone" and "Something." New York: Oxford University Press; 2006. 265 p.

32. Siedentop L. Inventing the Individual: The Origins of Western Liberalism. 1st ed. Cambridge, Mass.: Belknap Press; 2014. 448 p.
33. Spaemann R. Human Dignity. In: De Graaff G, editor. Essays in Anthropology: Variations on a Theme. Eugene, Ore.: Cascade Books; 2010. p. 49–72.
34. Rist JM. What is a Person?: Realities, Constructs, Illusions. 1st ed. New York: Cambridge University Press; 2020.
35. Hurtado LW. Honoring the Son: Jesus in Earliest Christian Devotional Practice. Bellingham: Lexham Press; 2018. 96 p. (Bird MF, editor. Snapshots).
36. Bauckham R. Jesus and the God of Israel: God Crucified and Other Studies on the New Testament's Christology of Divine Identity. Grand Rapids, Mich.: Eerdmans; 2008. 336 p.
37. Augustine of Hippo. Teaching Christianity [De doctrina Christiana] [396-426]. 1st ed. Hyde Park, N.Y.: New City Press; 1996. 274 p. (Rotelle JE, editor. WSA).
38. Gregory I. Moralia in Iob; Commento Morale a Giobbe 3 (XIX-XXVII) [586–590]. Siniscalco P, editor. Rome: Città Nuova; 1997. (Opere di Gregorio Magno).
39. Wales JJ. Contemplative Compassion: Gregory the Great's Development of Augustine's Views on Love of Neighbor and Likeness to God. Augustin Stud. 2018 Fall;49(2):199–219.
40. Dennett DC. Intentional Systems Theory. Oxf Handb Philos Mind [Internet]. 2009 [cited 2018 Oct 2]; Available from: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199262618.001.0001/oxfordhb-9780199262618-e-20
41. Dennett DC. The Self as a Center of Narrative Gravity. In: Kessel F, Cole P, Johnson D, editors. Self and Consciousness: Multiple Perspectives [Internet]. Mahwah, N.J.: Erlbaum; 1992 [cited 2018 Nov 5]. Available from: http://cogprints.org/266/1/selfctr.htm
42. Bryson JJ. Robots Should Be Slaves. In: Wilks Y, editor. Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues. Philadelphia, Pa.: John Benjamins Publishing Company; 2010. p. 63–74. (Natural Language Processing).
43. Redstone J. Making Sense of Empathy with Social Robots. In: Nørskov M, editor. Social Robots: Boundaries, Potential Challenges [Internet]. Burlington, Vt.: Ashgate; 2016. p. 19–38. Available from: http://dx.doi.org/10.3233/978-1-61499-480-0-171
44. Bendel O. The Morality Menu [Internet]. 2019 [cited 2020 Aug 21]. Available from: https://maschinenethik.net/wp-content/uploads/2019/12/Bendel_MOME_2019.pdf
45. Augustine of Hippo. The City of God, Against the Pagans [413–427]. In: St Augustine's City of God and Christian Doctrine. Buffalo, N.Y.: Christian Literature Publishing Co.; 1887. (Schaff P, editor. Nicene and Post-Nicene Fathers, First Series).
46. Augustine of Hippo. The Literal Meaning of Genesis [De Genesi ad litteram] [401-415]. In: On Genesis. Hyde Park, N.Y.: New City Press; 2004. p. 168–506. (WSA).
47. Augustine of Hippo. The Trinity [399-419]. 1st ed. Hyde Park, N.Y.: New City Press; 1991. 472 p. (Rotelle JE, editor. WSA).
48. Turkle S. Alone Together: Why We Expect More from Technology and Less from Each Other. Revised and Expanded Ed. Basic Books; 2017. 400 p.
49. Sullins JP. Friends by Design: A Design Philosophy for Personal Robotics Technology. In: Kroes P, Vermaas PE, Light A, Moore SA, editors. Philosophy and Design: From Engineering to Architecture [Internet]. Dordrecht: Springer Netherlands; 2008 [cited 2019 Sep 6]. p. 143–57. Available from: https://doi.org/10.1007/978-1-4020-6591-0_11
50. Najork M. Using Machine Learning to Improve the Email Experience. In: Proceedings of the 25th ACM International Conference on Information and Knowledge Management. 2016. p. 891.
51. Loewenstein G, O'Donoghue T, Bhatia S. Modeling the interplay between affect and deliberation. Decision. 2015 Apr;2(2):55–81.
52. Troy AS, Wilhelm FH, Shallcross AJ, Mauss IB. Seeing the Silver Lining: Cognitive Reappraisal Ability Moderates the Relationship Between Stress and Depressive Symptoms. Emot Wash DC. 2010 Dec;10(6):783–95.