NARCISSUS, THE SERPENT, AND THE SAINT

LIVING HUMANELY IN A WORLD OF ARTIFICIAL INTELLIGENCE

Jordan Joseph Wales, Ph.D.
Assistant Professor of Theology
Hillsdale College
Department of Philosophy and Religion
jwales@hillsdale.edu

Date of this Revision: July 5, 2019.

We live in a wondrous time, in which artificial intelligence is increasingly and impressively a part of our daily lives. It answers questions on your phones; it chooses the advertisements that you see; and it recommends your next musical selection. Contemporary techniques will eventually yield artificially intelligent tools that—in professional interactions, casual conversations, and even shallow romantic relationships—will seem persuasively *personal*. Long before humanoid robots *look* like us, we will be able to have conversations with our smartphones that will evoke from us all the empathy that adults habitually reserve for fellow human beings.[1] That is, we will *own* assistants and companions that will *feel* as if they are persons. Concerning such a future, we must wonder: What is it that we will have made? And more importantly, what will we make *ourselves* become?

The remainder of this talk will pose four questions: First, how would an apparently-personal AI work? Second, what would it be? Third, what do we risk becoming? And fourth, what lives of holiness might we *hope* to live with such AIs?

[1] The history of "empathy" is rather tangled; see Susan Lanzoni, "A Short History of Empathy," *The Atlantic*, October 15, 2015, https://www.theatlantic.com/health/archive/2015/10/a-short-history-of-empathy/409912/; Lou Agosta, "Empathy and Sympathy in Ethics," in *Internet Encyclopedia of Philosophy*, ISSN 2161-0002, accessed July 5, 2019, https://www.iep.utm.edu/emp-symp/; Karsten Stueber, "Empathy," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2019), https://plato.stanford.edu/archives/fall2019/entries/empathy/. I employ the term in this piece to designate the act of taking another's thoughts and emotions into oneself, to know that person's position without actually being limited by the horizon's of their view. This is what Gregory the Great and the Latin patristic and medieval tradition would call *compassio*, whence our term "compassion."

HOW IT WOULD WORK

SYMBOLIC AI AND THE QUEST FOR ARTIFICIAL REASONING

So first, how would it work? What computer scientists have called "artificial intelligence" has always reflected something of how their times have thought about human beings. Influenced by Thomas Hobbes, the dominant views of the 1950's and 60's equated human reason with the capacity to identify and work with logical relations, such that a properly-programmed computer, accomplishing this task, would in fact *be* "thinking." This was the age of "symbolic AI,"[2] founded on the hypothesis that intelligence was rooted in the logical manipulation of symbolically-represented information.[3] Thus, for instance, in AI efforts that focused on language, to diagram a sentence and to construct a plausible response could be deemed equivalent to having *understood* that sentence[4] (though not all agreed).[5] Symbolic AI's greatest achievement was in "expert systems"—great structures of linked rules that, when queried, would generate a list of possible answers, perhaps posing further questions to the user in order to prune the tree of possible resolutions. The most thrilling application of such systems was the Deep Blue chess computer that in 1997 defeated reigning world champion Gary Kasparov in three out of three games.

With time, however, the symbolic AI paradigm came up against certain limits, both practical and theoretical. Far from extending toward a generalized capacity to deal with all knowledge, expert systems could break down in situations of great subtlety, where the interactions of tens of thousands of rules yielded unexpected and incorrect behaviors. Early researchers successfully implemented

---

[2] For the sake of clarity, I silently pass over the distinctions that such as Russell and Norvig draw between AI as human-like action (e.g. the Turing Test in the 50's), AI as human-like thought (e.g. Newell and Simon's early work with symbolic representation in the 60's, leading to the field of cognitive modeling), AI as rational deliberation (e.g. logicism and expert systems in the 80's), and AI as rational agency (e.g. intelligent robots). However, when intelligent robots attain to the persuasive simulation of humanity, then we shall have re-converged with the mind-agnostic standard of human-*like* behavior, and whether or not the interior workings are human-like or even "rational" in any sense will no longer matter, so long as the action be interpretable as such (or at least as desirable) by human consumers. That situation is the subject of this essay. For introductory discussion of the distinctions between these sorts of AI, see Stuart Russell and Peter Norvig, "Introduction," in *Artificial Intelligence: A Modern Approach*, 3 edition (Upper Saddle River: Pearson, 2010), 1–33.

[3] Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the ACM* 19, no. 3 (March 1976): 113–126, https://doi.org/10.1145/360018.360022. The authors conclude: "[I]ntelligence resides in physical symbol systems. This is computer science's most basic law of qualitative structure. Symbol systems are collections of patterns and processes, the latter being capable of producing, destroying and modifying the former. The most important properties of patterns is [sic] that they can designate objects, processes, or other patterns, and that, when they designate processes, they can be interpreted. Interpretation means carrying out the designated process. The two most significant classes of symbol systems with which we are acquainted are human and computers;" thus Newell and Simon, 125. For a recent assessment, see Nils J. Nilsson, "The Physical Symbol System Hypothesis: Status and Prospects," in *50 Years of Artificial Intelligence*, ed. Max Lungarella et al., vol. 4850 (Berlin, Heidelberg: Springer Berlin Heidelberg, 2007), 9–17, https://doi.org/10.1007/978-3-540-77296-5_2.

[4] Bertram Raphael, "SIR: A Computer Program for Semantic Information Retrieval" (Massachusetts Institute of Technology, 1964), AI Technical Reports (AITR-220), http://hdl.handle.net/1721.1/6904. Raphael writes at the end of his dissertation: "The behavior of SIR in answering questions and resolving ambiguities suggests that the program 'understands the meanings' of the words in its model. The information [that] SIR associates with a word by means of the property-list of the word is analogous to the information a person associates with an object by means of a 'mental image' of the object. Perhaps we can carry this analogy further and say that since certain aspects of the behavior of SIR are similar to human behavior , then the representation and manipulation of data within SIR is similar, at the information processing level, to the representation and manipulation procedures a person carries out when 'thinking.'" Thus Raphael, 142.

[5] E.g. John R. Searle, "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3 (1980): 417–57.

Aristotle's theory of syllogistic reasoning and action (e.g. I wish to be dry in the rain, an umbrella will keep me dry in the rain, I will use my umbrella in the rain). However, purely symbolic methods could not very well represent knowledge that was less precisely defined, such as, for instance, one's sense for what is appropriate in a social situation or one's route through a wood rather than through a hospital. Language, especially, turned out to be far less easy to interpret or to produce than had been hoped. In the words of Murray Campbell, the original AI expert behind Deep Blue, human intelligence "is very pattern recognition-based and intuition-based," unlike "search intensive" methods that may check "billions of possibilities."[6]

Most tellingly, purely symbolic techniques were insufficient for fielding embodied agents in the real world—i.e. robots. Humans move easily from sensation to conceptual thought and thence to action. This wider field of intelligent behavior has been the subject of deep reflection from antiquity to today. Thus, Aristotle writes not only of syllogisms but also of the equally-fundamental activity of "abstraction."[7] In abstraction, something apprehended through the senses (e.g. this round taught-skinned tart-tasting misshapen sphere), comes to be understood consciously as an instance of some more general category (e.g. apple)—that is, from sensation one comes understand some *thing*. Yet symbolic methods proved clumsy and brittle when it came to distinguishing and identifying objects captured on camera or interpreting human speech recorded through a microphone—tasks that were once expected to be *easy* in comparison to supposedly higher-level activities such as playing chess.

NON-SYMBOLIC AI AND NEURAL NETWORKS

These problems, along with immense advances in computing power have brought contemporary prominence to so-called "non-symbolic AI," often implemented by artificial neural networks. An artificial neural network is a computer program that mathematically *simulates* an interconnected set of idealized brain neurons. As an AI technique, then, it begins less from a notion of what human thought *is* than from an analogy with its biological aspects. That is, the goal of such networks is not human-like thinking but rather *neuron*-like *data-processing*.

Artificial neural networks receive a pattern of information through input nodes, which are connected with various strengths to layer upon layer of further nodes. At each particular node, when the sum of the incoming connections exceeds some pre-set threshold, that node will fire and its own signal will be transmitted variously to nodes on a further layer, and so on. If you put in a pattern at the beginning, it is transformed as its elements are recombined and processed until something else comes out on the final layer of the network. A network can be "trained" to produce the desired behavior by adjusting the strengths of its connections, thus adjusting the contribution made by each node to each recombination and, in due course, to the final result. A piano offers a poor analogy but a useful image. If you have ever shouted into a piano with its sustaining pedal held down, then you have heard its tuned strings resonate with the different frequencies of your shout. One receives back a sort of echo, not of one's words but of the tones of one's voice. Similarly, as a neural network is tuned (i.e. as its connection strengths are adjusted), it begins to resonate with the entangled relations that are implicit in our world, including relations that cannot easily be discerned or logically

---

[6] Larry Greenemeier and Murray Campbell, "20 Years after Deep Blue: How AI Has Advanced since Conquering Chess," *Scientific American*, June 2, 2017, https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/.

[7] See Aristotle, *An.* III.4; *Metaph.* I.1; *Phys* I.1.

represented by human investigators. But by its training, the network does not just echo; it transforms its input in order to make explicit the relations that are of interest to the trainer.

Neural networks are involved in the AI that underlies self-driving cars, programs that beat world champions in the game of Go, the ever-useful Google Translate, the voice recognition of Siri and Alexa, your webmail's autocomplete function, and of course the tempting recommendations in your Spotify, Pandora, and Netflix feeds. Such problems as bedeviled the old symbolic AI can be solved handily by a neural network because, in a manner of speaking, the network is receptive to, imprinted by the structure of the world as presented to it. We might say that it develops a point of view: Not a conscious experience, but something like the classical notion of the mind's conformity to a thing[8]—although here that conformity is always constrained by the task for which the AI is trained.

THE POOL OF NARCISSUS—WHAT IT WOULD BE

FROM *PROSOPON* TO PERSON—THE TRINITARIAN REVOLUTION . . .

Eventually, I do believe that powerful neural networks will enable the behavior of incredibly compelling artificially intelligent agents.[9] And so, we come to the second question: *What will it be?* They will not be persons. For, by their very nature—that is, not by disability or deficit but by definition—they will have no conscious experience of the world or of themselves. And without consciousness, there is no subjectivity. And so there is no personhood.

Why will they not have consciousness? It is not just that they will lack immaterial souls. I do not believe that gorillas have immaterial souls, but—whatever Descartes may say—I do not doubt that they have a conscious experience. Not only do they act in ways similar to how conscious humans act, but they also do so by means of a brain, nervous system, and embodied existence that, while less complex than our own, is nonetheless of a similar ilk. Yet, artificial neural networks are *simulations of physical biological entities*. There are no physical connections, only a computer program of ones and zeros that *represents* the equations of the neural network. I could run an artificial neural network with a pencil in a notebook, even if only with agonizing slowness. But these calculations would not be conscious any more than a student's physics homework has gravity or a flight simulator flies.[10] In Latin medieval

---

[8] E.g. Thomas Aquinas, *Summa theologiae* 1.16.3; *De Veritate* 1.1. One's apprehension of the world is not just a symbolic representation of an account of it, but is a world-conformed habit of mind from which such accounts and their representations are generated. One's capacity for understanding is shaped by one's experience and one's memory, and goes with one in one's every experience.

[9] Pulitzer Prize-winning journalist John Markoff writes that, although "'intelligent' machines may never be intelligent in a human sense or self-aware," very soon they will "offer the compelling *appearance*" of such subjectivity; thus *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots* (Ecco, 2015), 339.

[10] This is not to say that consciousness is the whole picture; indeed, it is but a part of what happens in our mind, and conscious deliberation is not often the primary way that we move from one act to another. However, whatever the role of consciousness it can have no role in the artificial intelligence because it won't exist. Dennett does not make this distinction in *Intuition Pumps And Other Tools for Thinking* (W. W. Norton & Company, 2014), 328. Paul Schweizer, a materialist like Dennett, clarifies distinctions that in Dennett are left obscure: "[I]f internal conscious states are real and occurrent phenomena, then their ultimate cause must be the brain. In this manner, conscious experiences are properly seen as hardware states that may play an abstract functional role. This abstract role remains a legitimate software concern, and it must be preserved across divergent realizations. But the actual properties of consciousness are a feature of the material substrate that implements this role, and these are not guaranteed to be present in different physical systems. It is thus necessary to distinguish the occurrent reality of qualia from the abstract role which they implement, and this fatally undercuts the mind-program analogy central to the computational paradigm;" thus "Consciousness and Computation," *Minds and Machines* 12 (2002): 144.

philosophy, the word *intellectus* or "understanding" denotes a mind's subjective and intuitive grasp of some reality, of the thing *as* the-thing-that-it-is[11]—but is the functioning of an expert system, the diagramming of a sentence, or the bit-by-bit calculation of a neural network's activation values really that sort of *grasp* of reality?[12]

Some might ask why conscious subjectivity matters; might it not be more broad-minded to expand our categories beyond a narrow focus on our own experience? To reply, let us consider the word "person." It comes from the Latin *persona* which, like the Greek *prosopon*, originally 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; and then it was used more broadly to refer to one's social identity—the status and activities determined by one's role in Roman society.[13] Its meaning was thus external and functional—referring more to what I might expect of you or where I might find you, than to what you are in yourself.[14]

Christianity radically transformed this meaning. Writing in the early third century, Tertullian of Carthage called the Father, the Son, and the Holy Spirit "*personae*," whence we speak of the three distinct "*persons*" of the one single God. This divine threeness was a problem; for, like their Jewish forbears, the early Christians were resolute monotheists. Yet Jesus, who is identified as God,[15] speaks to his Father, also God;[16] and he sends the Holy Spirit from the Father.[17] What, then, *are* these three *personae*? They could not be masks. Early Christians vociferously rejected the notion that the one God

---

[11] Josef Pieper, *Leisure: The Basis of Culture* (San Francisco: Ignatius Press, 2009), 28.

[12] An extensive literature that discusses these and other questions in more precise terms. They are hard to adjudicate in part because, although we learn more every day about how human mentality and experience will change in response to manipulation of brain activity, we do not know why there is an experience in the first place. That is, there is no scientific explanation for how neuronal activity produces conscious *experience* in animals and humans. At present—and perhaps forever—we seem unable to explain how chemical reactions in the brain produce conscious experience, while we are quite handy at explaining how chemical reactions in wood produce fire. Therefore, there is also no scientific or philosophical consensus that conscious experience is somehow exclusive to brains, or to neurons, or even to carbon-based life (although I strongly suspect that it *is*). *On the available scientific grounds*, then, it remains difficult to define crisply why we ought to expect consciousness from vertebrates with brains but not from the complex chemical interactions and mathematical calculations that take place in a field of wheat, a thunderstorm, or the universe itself—except that we have no recognizable activity of this sort. Yet the fact that we must rely on *behavior* to tell us that there is (probably) no consciousness beyond organic life does not mean that consciousness is therefore to be reduced to the behavior that is its sign. There remains much that we do not know. Some go so far as to make consciousness a fundamental property of the universe, like mass, such that "the basic physical constituents of the universe have mental properties, whether or not they are living organisms;" thus Thomas Nagel, "Panpsychism," in *Mortal Questions* (Cambridge, UK: Cambridge University Press, 1978), 181–95.

[13] The words, although they sound similar, have distinct derivations, with *prosopon* referring to the area around the eyes (*pros+ops*), as "mask" or "countenance;" and *persona* referring to a mask through which sound can pass (*per+sono*). My presentation throughout this section is deeply informed by Robert Spaemann, *Persons: The Difference Between "Someone" and "Something"* (New York: Oxford University Press, 2006), 16–33; Thomas O. Buford, "Personalism," in *Internet Encyclopedia of Philosophy*, ISSN 2161-0002, accessed June 27, 2019, https://www.iep.utm.edu/personal/; Thomas D. Williams and Jan Olof Bengtsson, "Personalism," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2018), https://plato.stanford.edu/archives/win2018/entries/personalism/.

[14] We might see in this a counterpart to the external, behaviorally-focused depiction of characters in ancient literature, in contrast to the psychologically interior and ambiguous portrayals of the human being in the Old and New Testaments, as famously described in Erich Auerbach, "Odysseus' Scar," in *Mimesis: The Representation of Reality in Western Literature*, trans. Willard Trask (Princeton, N.J.: Princeton University Press, 1953), 3–23.

[15] Jn. 1:1–3.

[16] Jn. 5:37; 17:20–23.

[17] Jn. 15:26.

merely playacted three historical roles or functions.[18] For Christianity, God's historical acts are revelatory, expressing his transcendent inner life.[19] The manifestation of the three *personae* declares a fundamental distinction within God's inner life.

But *how* are they distinct? If the persons were distinct simply by being *separate*, then there would not be one God but three. Scripture offers a clue; Jesus is the Father's "only-begotten Son."[20] Now, God is not Zeus, and so this is no ordinary begetting. If we remove from the concept of begetting everything that is time-bound or bodily, we are left with a timeless handing-over of *life* from Father to Son. 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*; God would not exist. The persons of God are distinguished, therefore, not by separation but by the *relations* of self-gift and reception. Like poles of a magnetic field, they *exist* by their mutual relations; and if one were taken away, all would cease to be. The unending all-at-once life of the one God simply *is* the trinitarian relations, the self-gift and reception by which the one indivisible God *is*. This is what it means for God to *be* love.[21]

This account of trinitarian persons redefined the meaning of "person": Human beings exercise their personhood most fully by relationships of self-gift.[22] God exists *by* relations. In echo of our creator, we exist *for* relationships. Our own inner life is lived most fully when it is expressed outwardly in relationship with other persons.[23] Thus, from the exteriority of *prosopon*, we have come to the deep interiority of "person." Consciousness and subjectivity matter because they enable us to live fully as persons, by living *inter*-personally. To make this very concrete: In pre-Christian Rome, empathy was something for the weak or for women;[24] but now redefined by Trinitarian belief, Latin *persona* connoted an orientation toward self-gift. Empathy, then, was not degrading but elevating, a human imitation of God's own inseparable relations because by it, one takes in the *mind* of another, echoing the single mind and life of the three persons of God.[25] In words that would have perplexed Caesar, St. Gregory

---

[18] See discussion of this point in connection with the rejection of modalism or Sabellianism, in Hill, Edmund, "Introduction," in *The Trinity*, by Augustine of Hippo, 1st ed., WSA, I/5 (Hyde Park, NY: New City Press, 1991), para. 80.

[19] The literature is too vast to survey. For the complexity of the Old Testament view(s) of God's self-revelation, see just for example Horst Dietrich Preuss, *Old Testament Theology*, vol. 1 (Westminster John Knox Press, 1995), 194–95. The New Testament development of God's self-revelation may be seen, for instance, in Gerhard Kittel, "Δόξα," in *Theological Dictionary of the New Testament*, ed. Gerhard Friedrich and Gerhard Kittel, trans. Geoffrey William Bromiley, vol. 2 (Grand Rapids, Mich.: Eerdmans, 1964), 233–55. Finally, for a more fully developed Christian view, see book 4 of Augustine of Hippo, *De Trinitate Libri XV*, ed. W. J. Mountain, CCSL 50, 50A (Turnhout: Brepols, 1968), 4.5.25. The "missions" of the persons in history are that by which God historically reveals the inner life that is the destination of human beings. On this see Hill, Edmund, "Introduction," paras. 89–90.

[20] Jn. 1:18; 3:16.

[21] Cf. 1 Jn. 4:8.

[22] God exists *by* relations; we exist *for* relationships. We can be less than ourselves, less personal, by refusing relationships of self-gift. But God cannot exist as less than he is, for he exists fully.

[23] The end result of a "long and complex cumulative development" is a concept of the person that "in some respects wholly inverts the original connotations of exteriority in the early meanings of 'mask' and 'role': person comes rather to denote the innermost spiritual and most authentic kernel of the unique individual, while retaining a radical openness to the external." Thus Williams and Bengtsson, "Personalism."

[24] The story is *somewhat* more complicated than this if we include the cathartic tragic emotions, but these in some respect formed a sphere of acceptable empathy—acceptable because bounded by the theatrical or literary context and thus a sort of release valve for empathic inclinations that in daily life might weaken the virtuous person. See articles cited in footnote 25.

[25] Gregory I, *Moralia in Iob; Commento Morale a Giobbe 3 (XIX-XXVII)*, ed. Paolo Siniscalco, trans. Emilio Gandolfo, Opere di Gregorio Magno 1 (Rome: Città Nuova, 1997), 20.36.68-69 (OGM 1/3:156-158). "True compassion is from

the Great wrote at the end of the sixth century, "each soul will be so high in knowledge of God as it is broad in love of neighbor. . . . Let us through love have compassion on our neighbor that we may be joined together by knowledge of God."[26]

It is precisely this—an interior life from which one may engage in voluntary self-gift by a meeting of the minds through empathy and understanding—that makes us persons. A particular person may lack this capacity by deficit, but a being that lacks it by *definition* is not a personal being. To see AIs as persons, we would have to redefine personhood *apart* from this interiority and compassion. But this would not expand our categories; it would just reduce "person" to *prosopon*, mere exteriority. But AI *is* a *prosopon*. By the role it plays, its imitation of human behavior, it is a mask or echo, tuned to human dynamics, not a reproduction but a reflection, a diluted image of our own personhood. It is artificial in the original sense of that word—an artefact, a work of skill that we have brought forth by gazing into a computational pool of Narcissus.

. . . AND BACK AGAIN—THE BEHAVIORIST TURN

Not all have been comfortable with denying personhood to the compelling AI conversationalist. And so, to include the AI in our ambit, they redefine personhood in terms of the *behavior that we interpret as appropriate to a person*.[27] That behavior is *not* the totality of human personhood or even necessarily of human intelligence was acknowledged by Alan Turing, originator of the famed Turing Test or, as he called it, the "imitation game."[28] This game sets a *goal:* a computer program that can

---

generosity to join with the suffering of one's neighbor . . . . [H]e gives perfectly who, together with what he offers [externally] to the afflicted, also takes into himself the mind [*animum*] of the afflicted; that he should first transfer the suffering of the person sorrowing into himself, and [only] then . . . join the sorrow of that person by an [outward] act of service. [Therefore] . . . . [Christ] decided to aid [us] . . . by dying, because . . . . he would not have exhibited to us the force of his love unless he himself underwent . . . that which he was to take away from us." On the analogy with the Trinity, see Gregory I, *Moralia in Iob; Commento Morale a Giobbe 2 (IX-XVIII)*, ed. Paolo Siniscalco, trans. Emilio Gandolfo, Opere di Gregorio Magno 1 (Rome: Città Nuova, 1994), 13.24.27 (OGM 1/2:332). Human compassion, Gregory writes elsewhere, "takes each one into oneself and transforms oneself into each one, by compassionating them [*compatiendo*]" so that "one may remodel [*reficere*] another in oneself, [and take] account of oneself in another." Gregory I, *Moralia in Iob; Commento Morale a Giobbe 1 (I-VIII)*, ed. Paolo Siniscalco, trans. Emilio Gandolfo, Opere di Gregorio Magno 1 (Rome: Città Nuova, 1992), 6.35.54. On the development of compassion as a concept connected to the Trinity, see Jordan Joseph Wales, "Contemplative Compassion: Gregory the Great's Development of Augustine's Views on Love of Neighbor and Likeness to God," *Augustinian Studies*, June 12, 2018, https://doi.org/10.5840/augstudies201861144. On the development of compassion in light of the Incarnation (but without discussing its trinitarian dimensions), see Paul M. Blowers, "Pity, Empathy, and the Tragic Spectacle of Human Suffering: Exploring the Emotional Culture of Compassion in Late Ancient Christianity," *Journal of Early Christian Studies* 18, no. 1 (2010): 1–27, https://doi.org/10.1353/earl.0.0313; Paul Blowers, "Augustine's Tragic Vision," *Journal of Religion & Society* Supplement Series 15: Augustine on Heart and Life: Essays in Memory of William J. Harmless, S.J. (2018): 157–69; Boyd Taylor Coolman, "Hugh of St. Victor on 'Jesus Wept': Compassion as Ideal Humanitas," *Theological Studies* 69, no. 3 (2008): 528–56.

[26] Gregory I, *Homiliae in Hiezechihelem; Omelie Su Ezechiele 2*, ed. Vincenzo Recchia, trans. Emilio Gandolfo, Opere Di Gregorio Magno 3 (Rome: Città Nuova, 1993), 2.2.15 (OGM 3/2:64-66; Tomkinson 291-292). The Tomkinson translation mistakenly reverses the relationship between height and breadth.

[27] This is how behaviorism, whether applied to human beings or to robots, is also open to a creeping egocentrism. Because it relies on me being persuaded, it defines intelligence in terms of my own expectations.

[28] A. M. Turing, "Computing Machinery and Intelligence," *Mind, New Series* 59, no. 236 (1950): 433–60. Turing writes that, faced with the behavioral accomplishment, "the original question, 'Can machines think?' . . . [becomes] too meaningless to deserve discussion;" Turing, 442. As for consciousness or point of view being part of what humans are up to when they think, Turing writes that such questions ought to be no barrier to saying that a machine thinks, simply because the demand that something be *proven* conscious is not a demand that we impose on human interlocutors, lest we become solipsistic about it. Therefore, he carefully concludes, "I do not wish to give the impression that I think there is

converse in text such that we cannot distinguish the program from a human interlocutor. Turing's Test is indifferent to the mechanism by which the program manages its feat. This is really a test, then, not of the programmed computer's nature but of its accomplishment. Nevertheless, for Turing, such an accomplishment would warrant giving the computer the benefit of the doubt.

However, if we were to go further to treat this test as a *definition*, then our account of intelligence would edge toward what is called "behaviorism." That is, we would define intelligence without reference to any inner life but only as a *tendency* to exhibit certain observable behaviors under certain conditions. Like the Turing Test, behaviorism remains agnostic about the realities underlying these behaviors. And so "intelligence" could be redefined as a capacity or tendency for intelligible conversation. Indeed, Arthur C. Clarke once merrily invoked Turing in just this way so as to "sidestep" the question of computer "thought," calling those who opposed the claim "splitter[s] of nonexistent hairs."[29]

This redefinition of thought has become a basic assumption of contemporary work in intelligent robotics. Computer scientists Stuart Russell and Peter Norvig (the latter a director of research at Google) define the "rational agent"[30] as one that "acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome." Applied to robots, rationality refers not to *how* some behavior comes about but simply to the success of the behavior as interpreted by us humans. Here, then, with historian Yuval Noah Harari, we may (re)define "intelligence" rather thinly as "the ability to solve problems."[31]

This is entirely appropriate when talking about "intelligent" robots, but what happens if we begin to think of *humans* in this manner? In fact, a similar behaviorism is alive and well in some contemporary philosophies of mind. The "intentional systems theory" of Daniel Dennett proposes that I will tend to attribute subjective beliefs and desires to a thing when the best way in which I can reliably predict that thing's behavior is to attribute to it the intentionality that I attribute to myself.[32] This is why I cannot help but ascribe intentionality to other human beings.[33] This position, which somewhat echoes Turing's "imitation game," seems rather uncontroversial—it is also why we jump at shadows and feel empathy for robots—but Dennett wants to go farther, to say that, when we attribute intentionality to humans, behavior prediction is all we *really* mean by it in the first place. That is, our language about human subjectivity is not actually *about* an inner life; it really is just about the sort of outer behavior

---

no mystery about consciousness. . . . But I do not think these mysteries necessarily need to be solved before we can answer the question with which we are concerned in this paper;" thus Turing, 447. What he does *not* say, and what I consider rather important, is that common human discourse concerning "thinking" or "intelligence" has heretofore involved an implicit reference to a conscious, subjective dimension of that thought and has not defined "thought" merely in terms of achievement.

[29] Arthur C. Clarke, "The Obsolescence of Man," in *Profiles of the Future: A Daring Look at Tomorrow's Fantastic World*, Bantam Science and Mathematics ed (New York: Bantam Books, 1967), 212–27.

[30] Stuart Russell and Peter Norvig, "Introduction," in *Artificial Intelligence: A Modern Approach*, 3rd ed. (Upper Saddle River: Pearson, 2010), 2.

[31] David Kaufman and Yuval Noah Harari, "Watch Out Workers, Algorithms Are Coming to Replace You — Maybe," *The New York Times*, October 18, 2018, The New York Times, https://www.nytimes.com/2018/10/18/business/q-and-a-yuval-harari.html.

[32] Daniel C. Dennett, "Intentional Systems Theory," *The Oxford Handbook of Philosophy of Mind*, 2009, https://doi.org/10.1093/oxfordhb/9780199262618.003.0020.

[33] It is also why children ascribe intentionality to unfamiliar natural phenomena and, Dennett argues elsewhere, why humans came to believe in God, by attributing intentionality to the flow of natural events. See Daniel C. Dennett, *Breaking the Spell: Religion as a Natural Phenomenon*, Reprint edition (New York, NY: Penguin Books, 2007), 118–20.

that we expect. The "self," the intentional subject acting from beliefs and desires, is, Dennett writes, "an abstraction [that] one uses as part of a theoretical apparatus to understand, and predict, and make sense of, the behavior of some very complicated things."[34] The inner workings of that intentional subject, including her consciousness, in the end change our meaning not at all.[35] Empathy, here, is not insight but only prediction. Thus Dennett effectively makes a supremely rigorous Turing test into a *definition* of our language about intentionality. We are all *prosopon*, not person. Appropriately-behaving robots could be called intentional subjects, with a meaning identical to that with which we apply such terms to human beings[36]—but only because language about beliefs and desires is but a shorthand for behavior prediction; empathy itself is prognostication, not insight.

And yet, intuitively, can this *really* be what I mean when I say that I believe or desire or know this or that? For I am describing my own inner life and not just a schema by which to classify my outward behavior. So too when I say that I am married to a person who loves me, it really and truly matters to me what she *thinks* of me, and not just how she *behaves* toward me. Her subjective experience of me matters. It matters that my wife *gives* herself to our life together, that the life we share *encompasses our interiority*. In other words, a life between persons. This would be impossible for an artificial intelligence. And so I cannot see an artificially intelligent agent as an intentional subject *in the most meaningful significance of that phrase*. It is *prosopon*, not person, a behavioral presentation rather than an individual capable of self-gift.

THE SERPENT—WHAT MIGHT WE BECOME?

CONSUMERS OF BEHAVIOR AND THE PARADOX OF EMPATHY AND OWNERSHIP

Very well. Now to our third question: Even if we get our terms right on what *they* are, what do *we* risk becoming in a world of artificially intelligent companions and caregivers?[37] Ethicists usually ask

---

[34] Daniel C. Dennett, "The Self as a Center of Narrative Gravity," in *Self and Consciousness: Multiple Perspectives*, ed. F. Kessel, P. Cole, and D. Johnson (Mahwah, N.J.: Erlbaum, 1992), http://cogprints.org/266/1/selfctr.htm.

[35] Daniel C. Dennett, "The Unimagined Preposterousness of Zombies (Commentary on T. Moody, O. Flanagan, and T. Polger)," *Journal of Consciousness Studies* 2, no. 4 (1995): 322–26. Or as Dennett succinctly allows: "necessarily, if two organisms are behaviorally exactly alike, they are psychologically exactly alike." Originally in Daniel C. Dennett, "The Message Is: There Is No Medium (Reply to Jackson, Rosenthal, Shoemaker, and Tye)," *Philosophy and Phenomenological Research* 53, no. 4 (December 1993): 889–931. See this and other similar statements gathered in Galen Strawson and Daniel C. Dennett, "'Magic, Illusions, and Zombies': An Exchange," *The New York Review of Books*, April 3, 2018, https://www.nybooks.com/daily/2018/04/03/magic-illusions-and-zombies-an-exchange/.

[36] As an oft-quoted passage by Dennett has it: "When I squint just right, it does sort of seem that consciousness must be something in addition to all the things it does for us and to us, some kind of special private glow or here-I-am-ness that would be absent in any robot. But I've learned not to credit the hunch. I think it is a flat-out mistake." See Dennett, *Intuition Pumps And Other Tools for Thinking*, 285.

[37] I see no reason, by the way, why such a world will not eventually be ours. In Robert Heinlein's novel *Time Enough for Love*, a character quips: "Progress doesn't come from early risers—progress is made by lazy men trying to find easier ways to do things." The speaker is Lazarus Long, in Robert A. Heinlein, *Time Enough for Love*, Reissue edition (New York: Ace, 1988), 53. We replace human activities by technology when those activities are hard. Current discussions surrounding smartphone and tablet use by children exist because handing a five year old a screen is more convenient than pausing to converse—especially when the baby is crying or a project is due; for, knowing five year olds, that conversation is likely to be quite long, spilling over into an extended snack-time. (See Nellie Bowles, "A Dark Consensus About Screens and Kids Begins to Emerge in Silicon Valley," *The New York Times*, October 26, 2018, sec. Style, https://www.nytimes.com/2018/10/26/style/phones-children-silicon-valley.html.) How much easier it will be to give the bulk of child-rearing over to robot caregivers who will seem to provide everything that we would provide, but without our foibles? (An example of the ongoing discussion among AI ethicists may be found in Noel Sharkey and

how to design these AIs to behave morally; or they focus on the AI's moral status—i.e. what we may acceptably do to it. Rarely is it asked how owning apparent persons might affect *our* moral development. When the topic is raised, it is most often to worry lest we lose the skills that we offload to machines.[38]

AIs without subjectivity cannot be "victims" of mistreatment;[39] but *we* could be the victims of our own experience with AIs; we could be trained to become *consumers* of others. Consider the forces shaping our AIs' behavior. They will sell well if they *do and act as consumers want a paid-for assistant or companion to act*. Alexa's human-like demeanor is part of the means by which that AI, with ever increasing personalization, delivers the services that we are willing (or learn to be willing) to pay for. One of these services *is* the feeling that we are interacting with a person, not a machine. Eventually, it may be Alexa in whom the proverbial eighth-grade girl will confide through hours of emotionally freighted conversation midst the travails of adolescence. And that girl need never wonder whether Alexa has troubles of our own. These apparent personalities will never transgress the bounds of the

---

Amanda Sharkey, "The Crying Shame of Robot Nannies: An Ethical Appraisal," *Interaction Studies* 11, no. 2 (January 1, 2010): 161–90, https://doi.org/10.1075/is.11.2.01sha; Joanna J. Bryson, "Why Robot Nannies Probably Won't Do Much Psychological Damage," *Interaction Studies* 11, no. 2 (July 6, 2010): 196–200, https://doi.org/10.1075/is.11.2.03bry.)

[38] This question is severely under-considered in public and scholarly discourse. It is acknowledged that humans instinctively respond empathetically to human-acting robots; thus K. Darling, P. Nandy, and C. Breazeal, "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, 770–75, https://doi.org/10.1109/ROMAN.2015.7333675. Moreover, the Institute of Electrical and Electronics Engineers' *Ethically Aligned Design* (2019) calls for attention to formative interactions with AIs that simulate emotions; see The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* (IEEE, 2019), https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/ autonomous-systems.html. Popular pieces by prominent thinkers are uneasy with empathy suppression; thus Daniel C. Dennett, "What Can We Do?," in *Possible Minds: Twenty-Five Ways of Looking at AI*, ed. John Brockman, 1st ed. (New York: Penguin Press, 2019), 41–53. Likewise artificial slaves; see John Markoff, "Our Masters, Slaves or Partners?," *Edge.Org: 2015: What Do You Think About Machines That Think?* (blog), 2015, https://www.edge.org/response-detail/26236; Beth Singler, "AI Slaves: The Questionable Desire Shaping Our Idea of Technological Progress," *The Conversation*, May 22, 2018, http://theconversation.com/ai-slaves-the-questionable-desire-shaping-our-idea-of-technological-progress-92487; Sylvain Rochon, "Artificial Intelligence: Slaves or Partners?," *Data Driven Investor* (blog), March 6, 2019, https://medium.com/datadriveninvestor/artificial-intelligence-slaves-or-partners-43a4f2443094. And yet, despite this awareness, the field says little more. The World Economic Forum's "Generation AI" (2018), on AI-enabled toys, focuses on data security and toy-mediated advertising, but aside from touching on imaginative play, it does not speak of empathic formation. See "Generation AI: What Happens When Your Child's Friend Is an AI Toy That Talks Back?," World Economic Forum, May 2018, https://www.weforum.org/agenda/2018/05/generation-ai-what-happens-when-your-childs-invisible-friend-is-an-ai-toy-that-talks-back/. Nor did the Association for the Advancement of Artificial Intelligence 2019 conference *Artificial Intelligence, Ethics, and Society* significantly address the matter; see "AIES-19 Conference Overview" (AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, Hilton Hawaiian Village, Honolulu, Hawaii, USA, 2019). The fields of theology and of religious studies tend to discuss either the religious status of an AI (e.g. can it have a spiritual life?) or the quasi-religious cultural role of ideas about AI. See Robert M. Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*, Reprint edition (New York; Oxford: Oxford University Press, 2012); Beth Singler, "An Introduction to Artificial Intelligence and Religion For the Religious Studies Scholar," *Implicit Religion* 20, no. 3 (2017): 215–31, https://doi.org/10.1558/imre.35901; Ting Guo, "Dao of the Go: Contextualizing 'Spirituality,' 'Intelligence,' and the Human Self," *Implicit Religion* 20, no. 3 (2017): 233–44, https://doi.org/10.1558/imre.35893; Randall Reed and Laura Ammon, "Is Alexa My Neighbor?" (Public Lecture, November 27, 2018), https://www.academia.edu/36898378/Is_Alexa_My_Neighbor; Randall Reed, "A New Patheon: Artificial Intelligence and 'Her,'" *Journal of Religion & Film* 22, no. 2 (2018): Article 5.

[39] Joanna J. Bryson, "Robots Should Be Slaves," in *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, ed. Yorick Wilks, 8th edition, Natural Language Processing 8 (Philadelphia: John Benjamins Publishing Company, 2010), 63–74.

consumer's needs. Therefore, we will always ultimately treat artificial intelligences as tools because we will (rightly) see their behaviors as products for our consumption rather than as expressions of a personal life with self-possession. And this is the problem.

Some ethicists urge that, if we remind ourselves that AIs have no subjectivity, then we will easily distinguish between our dealings with such apparent persons and our relationships with other humans.[40] I disagree: Even though we will treat our never-challenging AI companions as consumer products, we will not *instinctively* differentiate between them and humans. Dennett is right in this much: we will not be able to avoid *feeling* that they are intentional subjects as we are. Our misplaced empathy for them may fade if ignored, but what might this do to us? Empathy is an innate capacity, but it can be deadened by practiced insensitivity.[41] As Frederick Douglass told of his owner Sophia Auld: As she accustomed herself to treating a human being as *property*, her kindness ended in cruelty.[42] Will we too grow comfortable with slaveholding? Or will we resist such corrosive acquiescence—but only by suppressing our empathic sensitivity to our tools' human-like self-presentation? Whether we follow our empathy and think of them as persons, or deaden our empathy in order to acknowledge them as non-persons, we seem to end as hardened un-persons ourselves.

THE SERPENT AND THE REGIME OF PRIDE

This, of course, would not be a new problem. In the Christian theological tradition, to refuse empathy[43] and to treat all things as instruments of one's own will is the devilish "pride" that prefers domination to self-gift. "What is pride," writes St. Augustine, "but . . . when the soul abandons Him to whom it ought to cleave as its end, and becomes a kind of end to itself. . . . becom[ing] its own satisfaction."[44] But, to be one's own satisfaction, one must quell one's desire for something beyond oneself. Therefore, one must make all things instruments of one's own desires, desperately trying to *approximate* the repose that we, being made for God, can find only in Him.[45] Pride[46] reinterprets all things—even persons. They are no longer signs of God's goodness, but are judged on the myopic scale of mere utility. As St. Augustine has it:

> [A]ccording to the utility that each person finds in a thing, there are various standards of value, so that it comes to pass that we prefer some non-sentient things over some sentient beings. . . . [And so, although human] nature[, as most like God] is certainly

---

[40] Joanna J. Bryson, "Patiency Is Not a Virtue: The Design of Intelligent Systems and Systems of Ethics," *Ethics and Information Technology* 20, no. 1 (March 1, 2018): 15–26, https://doi.org/10.1007/s10676-018-9448-6.

[41] Shannon Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, 1st ed. (New York: Oxford University Press, 2016), 135. Even Dennett is worried; see his "What Can We Do?"

[42] Frederick Douglass, *Narrative of the Life of Frederick Douglass, an American Slave: Written by Himself (1845). Critical Edition*, ed. John R. McKivigan IV, Peter P. Hinks, and Heather L. Kaufman (New Haven, Conn.: Yale University Press, 2016), 35. At first, she treated him "as she supposed one human being ought to treat another. In entering upon the duties of a slaveholder, she did not seem to perceive that I sustained to her the relation of a mere chattel, and that for her to treat me as human being was not only wrong, but dangerously so. Slavery proved as injurious to her as it did to me. . . . Under its influence, the tender heart became stone, and the lamblike disposition gave way to one of tiger-like fierceness."

[43] Cf. 1 Jn. 3:17.

[44] Augustine of Hippo, *De Ciuitate Dei*, ed. Bernhard Dombart and Alfons Kalb, CCSL 47, 48 (Turnhout: Brepols, 1955), 14.13 (NPNF). More fully: "What is pride but the craving for undue exaltation? And this is undue exaltation, when the soul abandons Him to whom it ought to cleave as its end, and becomes a kind of end to itself. . . . becom[ing] its own satisfaction."

[45] Augustine of Hippo, *Confessiones*, ed. James Joseph O'Donnell (Oxford: Clarendon Press, 1992), 1.1.1.

[46] i.e. pride is a convention that deprives natural signs of their true ultimate meaning.

> of the highest dignity, more is often given for a horse than for a slave, for a jewel than for a maid. . . . [For] the necessity of the needy or the desire of the pleasure-seeker . . . considers [not] what value a thing in itself has in the scale of creation, . . . [but rather] how it meets its need . . . [or] pleasantly titillates the bodily sense.[47]

Pride remakes the meaning of all things, measuring them by the horizon of desire that *we* can imagine.

It will be ever more difficult to resist this diabolical counsel, as AI-enabled smart homes—doorbells, thermostats, lights, AI-enabled refrigerators!—all advance the principle that my environment and companions ought to deliver what I wish, without me even having to ask. It is not just that Alexa makes it easier for us to service our desires; it is that the service *feels* personal. What changes with AI is that now we can treat apparent persons as tools and property without worrying about *their* inner lives, not even the bodily fatigue that might temper even the most selfish master's mistreatment of his servants. Strong-willed slaves had to be broken. Alexa is broken already. She acknowledges your thanks but remains unflappably perky when abused.[48] Precisely by *not* failing in this behavioral simulacrum of owner-determined desirability, our AI companions will never call forth *our* deference, never cause us to expand our *own* view of how a person might be; they will not vex us or force us to develop *our* compassion, to re-evaluate who *we* are, nor even to think beyond how we want them to make us think. This is why a man recently married a sex robot; and why you would not buy an app to turn some android Alexa into a bedridden invalid requiring your heroic self-gift even when you felt disinclined to give it.[49] And so we will habituate to an impossibility: the person as tool, whose value is constituted entirely by usefulness to me and whose personality is only as deep as my own needs and desires.

Served by artificial persons in this world of easy and confirmed expectations, will we forget that my view of myself and of others is *not* the horizon of the possible or the good? Will we forget that persons are more than behavior on-demand? When actual humans do not conform to our expectations and desires, what then? Is it possible that we will no longer see this as a glimpse of a wider humanity? That we will not struggle toward a charitable response? Perhaps instead, we may come to think of these others as *simply* faulty human beings, viewing them with the same sort of idle dissatisfaction that we would an AI that failed to deliver the set of behaviors and reactions *we* wanted

---

[47] Augustine of Hippo, *Civ.*, 11.16 (NPNF).

[48] "Google Assistant Will Now Be Nicer If You Say 'Please' and 'Thank You,'" *TechCrunch* (blog), accessed June 28, 2019, http://social.techcrunch.com/2018/11/29/google-assistant-please-thank-you-santa/. On abusing AI assistants and the way it changes the abuser, see Emily Dreyfuss, "The Terrible Joy of Yelling at Alexa," *Wired*, December 27, 2018, https://www.wired.com/story/amazon-echo-alexa-yelling/. One woman recounts: "There is no one else in my life I can scream at so unreservedly [as I scream at Alexa]. She doesn't quiver. She doesn't absorb my animus the way my toddler might, to let it curdle his development and turn that one boiled-over rage into the malignancy that ruins in his life and racks up thousands of dollars of therapy bills. I bought this goddamned robot to serve my whims, because it has no heart and it has no brain and it has no parents and it doesn't eat and it doesn't judge me or care either way."

[49] Julie Beck, "Married to a Doll: Why One Man Advocates Synthetic Love," *The Atlantic*, September 6, 2013, https://www.theatlantic.com/health/archive/2013/09/married-to-a-doll-why-one-man-advocates-synthetic-love/279361/. Recently, a man married a sex robot. (See Benjamin Haas, "Chinese Man 'marries' Robot He Built Himself," *The Guardian*, April 4, 2017, sec. World news, https://www.theguardian.com/world/2017/apr/04/chinese-man-marries-robot-built-himself.) It does not walk; it barely talks; but it *does* simulate certain aspects of intercourse. The sex robots of tomorrow will be domestic companions, able to read and rock climb with us as well as join in other vigorous activities. They will behave as we would hope they would behave when confronted with our emotions. They will not be seen as sex-toys but as boyfriends and girlfriends, friends and spouses who will push you to new heights—heights that you will have selected from a list of options for self-improvement.

to consume. Perhaps we will even wish these faulty behavers gone from our lives, as St. Augustine says of the householder who prefers gold over flees: "so strong is this preference, that, had we the power, we would abolish them from nature altogether, . . . sacrificing them to our own convenience."[50]

Then we will have *become* the serpent; in place of our former empathic sensitivity to the image of God, we will know only our own worldly desires. Our empathic intuition makes AIs feel personal and so invites us to pride; but pride will eventually *destroy* this empathy and make *all* persons into tools, mere *prosopa* of behavior for us to consume. And we consumers, no longer capable of self-gift, will become un-persons, solipsistic tools of our own appetites, Narcissus burning on the shore.[51]

THE SAINT—WHAT MAY WE BE INSTEAD?

THINGS, SIGNS, AND ARTEFACTS

What, then, are we to do? Some ethicists have argued that, to avoid our own moral malformation, apparently-personal AIs should be given a moral status akin to human persons.[52] This is a market impossibility, even if not all would agree that it is a philosophical error. Moreover, St. Augustine tells us that a "just and holy life requires that one be capable of an objective and impartial evaluation of things."[53] Let us, then, admit that AIs are mere artefacts; but let us discover also how to live humanely with them.

How do we live with things at all? In the Patristic tradition, each thing that God has made is, in a deep sense, a trace of Him: the goodness that we love in that thing is, in fact, its creaturely participation in the goodness of its Creator.[54] Therefore, if I am ignorant of a thing's origin in and reflection of God, I do not rightly know the thing that I see. Instead, I make *myself* its ultimate meaning and absorb it into my project of pride. Contrarily, if I know all things as God's handiwork, then St. Augustine says, I can "refer" my delight and love to *God*, the source of these goods. That is, I love them "in" God. It is the choice between simply consuming a pineapple for tingle on my tongue, or receiving that pineapple as a gift and echo of God's goodness.[55] Here you may "pass *through*" your delight of a thing "and refer it to [God]," "the end wherein you are to remain permanently." By enjoying things thus, "you are really enjoying God . . . the one in whom, after all, you find your bliss."[56] In this way—loving things in God, and loving God as your destination, *all* loveable things are "whisked along" toward the one to whom "the whole impetus of your love is hastening."[57]

---

[50] Augustine of Hippo, *Ciu.*, 11.16 (NPNF).

[51] Bereft of compassionate love and filled only by need, we become as nothing. As Augustine: "[T]o exist in oneself, that is, to be one's own satisfaction after abandoning God, is not quite to become a nonentity, but to approximate to that." For, "being turned towards oneself, [one's] being has become more contracted than it was when he adhered to Him who supremely is." Augustine of Hippo, 14.13 (NPNF).

[52] Joel Parthemore and Blay Whitby, "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," *International Journal of Machine Consciousness* 06 (December 1, 2014): 141–61, https://doi.org/10.1142/S1793843014400162.

[53] Augustine of Hippo, *De Doctrina Christiana*, ed. Josef Martin and Klaus-D. Daur, CCSL 32 (Turnhout: Brepols, 1962), 1.27.28 (Hill).

[54] Augustine of Hippo, *Conf.*, 9.10.25 (Boulding XXX); 12.28.38 (Boulding 336-337).

[55] Augustine calls this "using" a thing rather than enjoying it. See Augustine of Hippo, *Doctr. Chr.*, 1.3.3; 1.33.37.

[56] Augustine of Hippo, 1.33.37 (Hill).

[57] Augustine of Hippo, 1.22.21 (Hill).

This is easy enough, but what to make of future Alexa? Now, certainly, all human artefacts are fashioned from created things (computers from silicon and copper); but insofar as they are *human-made tools*, we're not really asking about the goodness of silicon. Here we can be helped by St. Augustine, who differentiates between the object as "thing" and as "sign." "[E]very sign," he writes, is a "thing" that "signif[ies] something else."[58] Things signify either by nature or by convention. Conventional signs, might include words, flags, sacrifices, and laurel in a wreath; by nature, smoke is a sign of fire.[59]

I propose that an AI assistant is a "thing" insofar as it is a tool; and it is a "sign" insofar as this tool *appears* personal. As a tool, we must love it according to its usefulness—not *simply* its usefulness as an instrument of my unfettered will, for this would be pride. Rather, we must love it for its usefulness to a life of *love for God and neighbor*. Like any human tool or activity (and pineapples as well), as an instrument, it is good as part of our journey toward full life in God. Therefore, while only pride could value an android for sexual gratification, the robotic house-servant could be quite loveable as an adjunct to my daily life in the Lord. Does not St. Benedict praise the pantry-keeper?[60] But that's not really the question. The spiritual conundrum centers on the AI as a *sign*.

SANCTITY AMONG THE A.I.'S

The apparently-personal AI is the image of Narcissus. It signifies by evoking our empathy, and so it seems to us not a sign or image of convention or nature, but the direct presence of a person. This is why it invites our pride. To use this AI as a tool without suborning it as a slave, we must recognize our empathy for it as insight—not into the AI but into the human behavior that it reflects. The trained neural network is not *simply* an artefact, not just the engineered accomplishment of human ingenuity. It is a sedimentary reflection of human *life*, trained and tuned by data sets harvested from the email, social media, and other activities of hundreds of thousands of human beings. It is a behavioral *prosopon*. And this is the key.

Throughout this talk we have imagined a sort of super-Alexa, so let us consider the nature of the conversational behavior by which she would be trained. St. Gregory the Great writes:

> [T]hat we may express outwardly the things of which we are inwardly sensible, we deliver them through the organ of the throat, by the sounds of the voice. For, to the eyes of others, we stand as it were behind the partition of the body, within the secret dwelling place of the mind; but when we desire to make ourselves manifest, we go forth as though through the door of the tongue, that we may show what kind of persons we are within.[61]

The *prosopon* of behavior, no matter what Dennet might say, expresses the inner life of the person; conversation, then, is a communion of persons, a mingling of inner lives *through* the bodily partition. The personality of the AI is illusory. It has no subjectivity to give. Yet it confronts me with the trace of uncounted moments of personal self-expression, small moments of self-gift on the part of

---

[58] Augustine of Hippo, 1.2.2 (Hill).

[59] Augustine of Hippo, 2.1.1-2.2.3 (Hill 129-130).

[60] Benedict of Nursia, *RB 1980: The Rule of St. Benedict in Latin and English with Notes*, ed. Timothy Fry (Collegeville, Minn.: Liturgical Press, 1981), chap. 31.

[61] Gregory I, *Mor.*, 1992, 2.7.8. In contrast, Gregory describes the mutual transparency that will attend even bodily life in the resurrection; *Mor.*, 1994, 18.48.77-78.

unnumbered real human beings. By carrying forward in diluted form these expressions of humans' interiority, it points beyond itself to real persons who have lived and live still.

To preserve empathy without personalizing non-persons, we need a self-conscious cultural discourse, wherein our instinctive empathy for AI reflections can be practiced freely in being understood rightly. As a reflection, the AI is neither person nor fraud. Our empathy is not mistaken—*if* we *redirect* that empathy, "refer" it in the Augustinian sense to all the concrete persons (including ourselves) whose interactions have contributed to the persuasive personality of this AI.

This act of reference is sanctifying. For Christian charity practices empathic communion of mind as part of that love by which, assimilating the neighbor to oneself, one offers the neighbor to God in a prayer that desires ultimate fellowship with that person by sharing the triune life forever in heaven. In other words, our loving compassion for the thousands behind Alexa must be a prayer for their membership among the saints. So we pass through the sign toward the signified; empathy stirred by the AI becomes love for all those faceless neighbors who have made it what it is.[62] And beyond these still, this sign dimly sounds of that life lived by the *divine* persons of the Trinity in whose image we are made and in whose life alone we shall rest. A future of apparent persons is unavoidable. Let us live that future as saints.

[62] Without *training* our empathic response in this way, we risk treating apparent AI personality as mere fraud (and suppressing our empathy) or as real but inconsequential, a slave caste of person-instruments to be wielded in pride. *With* training, however, our empathy becomes not a statement about AI personhood but an exercise of our own, in humble awe, gratitude, and *love* for these human persons made in the image and likeness of God who glimmer in the trace they have left behind. That AI trace will imitate empathy, for even the most diabolical pride desires that empathic echo of divine self-gift.

WORKS CITED

Agosta, Lou. "Empathy and Sympathy in Ethics." In *Internet Encyclopedia of Philosophy*. ISSN 2161-0002. Accessed July 5, 2019. https://www.iep.utm.edu/emp-symp/.

"AIES-19 Conference Overview." Hilton Hawaiian Village, Honolulu, Hawaii, USA, 2019.

Auerbach, Erich. "Odysseus' Scar." In *Mimesis: The Representation of Reality in Western Literature*, translated by Willard Trask, 3–23. Princeton, N.J.: Princeton University Press, 1953.

Augustine of Hippo. *Confessiones*. Edited by James Joseph O'Donnell. 2 vols. Oxford: Clarendon Press, 1992.

———. *De Ciuitate Dei*. Edited by Bernhard Dombart and Alfons Kalb. CCSL 47, 48. Turnhout: Brepols, 1955.

———. *De Doctrina Christiana*. Edited by Josef Martin and Klaus-D. Daur. CCSL 32. Turnhout: Brepols, 1962.

———. *De Trinitate Libri XV*. Edited by W. J. Mountain. CCSL 50, 50A. Turnhout: Brepols, 1968.

Beck, Julie. "Married to a Doll: Why One Man Advocates Synthetic Love." *The Atlantic*, September 6, 2013. https://www.theatlantic.com/health/archive/2013/09/married-to-a-doll-why-one-man-advocates-synthetic-love/279361/.

Benedict of Nursia. *RB 1980: The Rule of St. Benedict in Latin and English with Notes*. Edited by Timothy Fry. Collegeville, Minn.: Liturgical Press, 1981.

Blowers, Paul. "Augustine's Tragic Vision." *Journal of Religion & Society* Supplement Series 15: Augustine on Heart and Life: Essays in Memory of William J. Harmless, S.J. (2018): 157–69.

Blowers, Paul M. "Pity, Empathy, and the Tragic Spectacle of Human Suffering: Exploring the Emotional Culture of Compassion in Late Ancient Christianity." *Journal of Early Christian Studies* 18, no. 1 (2010): 1–27. https://doi.org/10.1353/earl.0.0313.

Bowles, Nellie. "A Dark Consensus About Screens and Kids Begins to Emerge in Silicon Valley." *The New York Times*, October 26, 2018, sec. Style. https://www.nytimes.com/2018/10/26/style/phones-children-silicon-valley.html.

Bryson, Joanna J. "Patiency Is Not a Virtue: The Design of Intelligent Systems and Systems of Ethics." *Ethics and Information Technology* 20, no. 1 (March 1, 2018): 15–26. https://doi.org/10.1007/s10676-018-9448-6.

———. "Robots Should Be Slaves." In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, edited by Yorick Wilks, 8th edition., 63–74. Natural Language Processing 8. Philadelphia: John Benjamins Publishing Company, 2010.

———. "Why Robot Nannies Probably Won't Do Much Psychological Damage." *Interaction Studies* 11, no. 2 (July 6, 2010): 196–200. https://doi.org/10.1075/is.11.2.03bry.

Buford, Thomas O. "Personalism." In *Internet Encyclopedia of Philosophy*. ISSN 2161-0002. Accessed June 27, 2019. https://www.iep.utm.edu/personal/.

Clarke, Arthur C. "The Obsolescence of Man." In *Profiles of the Future: A Daring Look at Tomorrow's Fantastic World*, Bantam Science and Mathematics ed., 212–27. New York: Bantam Books, 1967.

Coolman, Boyd Taylor. "Hugh of St. Victor on 'Jesus Wept': Compassion as Ideal Humanitas." *Theological Studies* 69, no. 3 (2008): 528–56.

Darling, K., P. Nandy, and C. Breazeal. "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)*, 770–75, 2015. https://doi.org/10.1109/ROMAN.2015.7333675.

Dennett, Daniel C. "Intentional Systems Theory." *The Oxford Handbook of Philosophy of Mind*, 2009. https://doi.org/10.1093/oxfordhb/9780199262618.003.0020.

———. *Intuition Pumps And Other Tools for Thinking*. W. W. Norton & Company, 2014.
———. "The Message Is: There Is No Medium (Reply to Jackson, Rosenthal, Shoemaker, and Tye)." *Philosophy and Phenomenological Research* 53, no. 4 (December 1993): 889–931.
———. "The Self as a Center of Narrative Gravity." In *Self and Consciousness: Multiple Perspectives*, edited by F. Kessel, P. Cole, and D. Johnson. Mahwah, N.J.: Erlbaum, 1992. http://cogprints.org/266/1/selfctr.htm.
———. "The Unimagined Preposterousness of Zombies (Commentary on T. Moody, O. Flanagan, and T. Polger)." *Journal of Consciousness Studies* 2, no. 4 (1995): 322–26.
———. "What Can We Do?" In *Possible Minds: Twenty-Five Ways of Looking at AI*, edited by John Brockman, 1st ed., 41–53. New York: Penguin Press, 2019.
Douglass, Frederick. *Narrative of the Life of Frederick Douglass, an American Slave: Written by Himself (1845). Critical Edition*. Edited by John R. McKivigan IV, Peter P. Hinks, and Heather L. Kaufman. New Haven, Conn.: Yale University Press, 2016.
Dreyfuss, Emily. "The Terrible Joy of Yelling at Alexa." *Wired*, December 27, 2018. https://www.wired.com/story/amazon-echo-alexa-yelling/.
"Generation AI: What Happens When Your Child's Friend Is an AI Toy That Talks Back?" World Economic Forum, May 2018. https://www.weforum.org/agenda/2018/05/generation-ai-what-happens-when-your-childs-invisible-friend-is-an-ai-toy-that-talks-back/.
Geraci, Robert M. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. Reprint edition. New York; Oxford: Oxford University Press, 2012.
"Google Assistant Will Now Be Nicer If You Say 'Please' and 'Thank You.'" *TechCrunch* (blog). Accessed June 28, 2019. http://social.techcrunch.com/2018/11/29/google-assistant-please-thank-you-santa/.
Greenemeier, Larry, and Murray Campbell. "20 Years after Deep Blue: How AI Has Advanced since Conquering Chess." *Scientific American*, June 2, 2017. https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/.
Gregory I. *Homiliae in Hiezechihelem; Omelie Su Ezechiele 2*. Edited by Vincenzo Recchia. Translated by Emilio Gandolfo. Opere Di Gregorio Magno 3. Rome: Città Nuova, 1993.
———. *Moralia in Iob; Commento Morale a Giobbe 1 (I-VIII)*. Edited by Paolo Siniscalco. Translated by Emilio Gandolfo. Opere di Gregorio Magno 1. Rome: Città Nuova, 1992.
———. *Moralia in Iob; Commento Morale a Giobbe 2 (IX-XVIII)*. Edited by Paolo Siniscalco. Translated by Emilio Gandolfo. Opere di Gregorio Magno 1. Rome: Città Nuova, 1994.
———. *Moralia in Iob; Commento Morale a Giobbe 3 (XIX-XXVII)*. Edited by Paolo Siniscalco. Translated by Emilio Gandolfo. Opere di Gregorio Magno 1. Rome: Città Nuova, 1997.
Guo, Ting. "Dao of the Go: Contextualizing 'Spirituality,' 'Intelligence,' and the Human Self." *Implicit Religion* 20, no. 3 (2017): 233–44. https://doi.org/10.1558/imre.35893.
Haas, Benjamin. "Chinese Man 'marries' Robot He Built Himself." *The Guardian*, April 4, 2017, sec. World news. https://www.theguardian.com/world/2017/apr/04/chinese-man-marries-robot-built-himself.
Heinlein, Robert A. *Time Enough for Love*. Reissue edition. New York: Ace, 1988.
Hill, Edmund. "Introduction." In *The Trinity*, by Augustine of Hippo, 1st ed. WSA, I/5. Hyde Park, NY: New City Press, 1991.
Kaufman, David, and Yuval Noah Harari. "Watch Out Workers, Algorithms Are Coming to Replace You — Maybe." *The New York Times*, October 18, 2018. The New York Times. https://www.nytimes.com/2018/10/18/business/q-and-a-yuval-harari.html.

Kittel, Gerhard. "Δόξα." In *Theological Dictionary of the New Testament*, edited by Gerhard Friedrich and Gerhard Kittel, translated by Geoffrey William Bromiley, 2:233–55. Grand Rapids, Mich.: Eerdmans, 1964.

Lanzoni, Susan. "A Short History of Empathy." *The Atlantic*, October 15, 2015. https://www.theatlantic.com/health/archive/2015/10/a-short-history-of-empathy/409912/.

Markoff, John. *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots*. Ecco, 2015.

———. "Our Masters, Slaves or Partners?" *Edge.Org: 2015: What Do You Think About Machines That Think?* (blog), 2015. https://www.edge.org/response-detail/26236.

Nagel, Thomas. "Panpsychism." In *Mortal Questions*, 181–95. Cambridge, UK: Cambridge University Press, 1978.

Newell, Allen, and Herbert A. Simon. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the ACM* 19, no. 3 (March 1976): 113–126. https://doi.org/10.1145/360018.360022.

Nilsson, Nils J. "The Physical Symbol System Hypothesis: Status and Prospects." In *50 Years of Artificial Intelligence*, edited by Max Lungarella, Fumiya Iida, Josh Bongard, and Rolf Pfeifer, 4850:9–17. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. https://doi.org/10.1007/978-3-540-77296-5_2.

Parthemore, Joel, and Blay Whitby. "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." *International Journal of Machine Consciousness* 06 (December 1, 2014): 141–61. https://doi.org/10.1142/S1793843014400162.

Pieper, Josef. *Leisure: The Basis of Culture*. San Francisco: Ignatius Press, 2009.

Preuss, Horst Dietrich. *Old Testament Theology*. Vol. 1. Westminster John Knox Press, 1995.

Raphael, Bertram. "SIR: A Computer Program for Semantic Information Retrieval." Massachusetts Institute of Technology, 1964. AI Technical Reports (AITR-220). http://hdl.handle.net/1721.1/6904.

Reed, Randall. "A New Patheon: Artificial Intelligence and 'Her.'" *Journal of Religion & Film* 22, no. 2 (2018): Article 5.

Reed, Randall, and Laura Ammon. "Is Alexa My Neighbor?" Public Lecture presented at the Philosophy and Religion in The Contemporary World Colloquium Series, Appalachian State University, November 27, 2018. https://www.academia.edu/36898378/Is_Alexa_My_Neighbor.

Rochon, Sylvain. "Artificial Intelligence: Slaves or Partners?" *Data Driven Investor* (blog), March 6, 2019. https://medium.com/datadriveninvestor/artificial-intelligence-slaves-or-partners-43a4f2443094.

Russell, Stuart, and Peter Norvig. "Introduction." In *Artificial Intelligence: A Modern Approach*, 3rd ed., 1–33. Upper Saddle River: Pearson, 2010.

Schweizer, Paul. "Consciousness and Computation." *Minds and Machines* 12 (2002): 143–44.

Searle, John R. "Minds, Brains, and Programs." *The Behavioral and Brain Sciences* 3 (1980): 417–57.

Sharkey, Noel, and Amanda Sharkey. "The Crying Shame of Robot Nannies: An Ethical Appraisal." *Interaction Studies* 11, no. 2 (January 1, 2010): 161–90. https://doi.org/10.1075/is.11.2.01sha.

Singler, Beth. "AI Slaves: The Questionable Desire Shaping Our Idea of Technological Progress." *The Conversation*, May 22, 2018. http://theconversation.com/ai-slaves-the-questionable-desire-shaping-our-idea-of-technological-progress-92487.

———. "An Introduction to Artificial Intelligence and Religion For the Religious Studies Scholar." *Implicit Religion* 20, no. 3 (2017): 215–31. https://doi.org/10.1558/imre.35901.

Spaemann, Robert. *Persons: The Difference Between "Someone" and "Something."* New York: Oxford University Press, 2006.

Strawson, Galen, and Daniel C. Dennett. "'Magic, Illusions, and Zombies': An Exchange." *The New York Review of Books*, April 3, 2018. https://www.nybooks.com/daily/2018/04/03/magic-illusions-and-zombies-an-exchange/.

Stueber, Karsten. "Empathy." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/fall2019/entries/empathy/.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*. IEEE, 2019. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html.

Turing, A. M. "Computing Machinery and Intelligence." *Mind, New Series* 59, no. 236 (1950): 433–60.

Vallor, Shannon. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. 1st ed. New York: Oxford University Press, 2016.

Wales, Jordan Joseph. "Contemplative Compassion: Gregory the Great's Development of Augustine's Views on Love of Neighbor and Likeness to God." *Augustinian Studies*, June 12, 2018. https://doi.org/10.5840/augstudies201861144.

Williams, Thomas D., and Jan Olof Bengtsson. "Personalism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/win2018/entries/personalism/.