Science and Engineering Ethics (2020) 26:2527–2532
https://doi.org/10.1007/s11948-020-00247-9

COMMENTARY

# Shifting Perspectives

**David J. Gunkel**[1]

Published online: 6 July 2020


The essay "Blame-Laden Moral Rebukes and the Morally Competent Robot: A Confucian Ethical Perspective" offers readers an innovative and potentially useful shift of perspective in AI/robot ethics. As the author accurately recognizes, the vast majority of work in the field—efforts that have been, for better or worse, organized around European/Christian moral traditions (i.e. utilitarianism, deontologism, virtue ethics, etc.)—tend to focus on the individual moral agent and therefore get hung up on questions regarding qualifying criteria for moral agency and status, i.e. personhood, consciousness, sentience, empathy, etc. The Confucian perspective mobilized by this contribution shifts the focus from the internal properties of the individual moral entity to the "moral ecology" of the human–robot system and the role that moral correctives or rebukes play in the management of this complex and multifaceted arrangement. The operative question that guides the inquiry is not whether robots in general or an individual robot in particular can be a moral agent but whether and to what extent an interactive artifact contributes to the development of a flourishing moral ecology within the context of human–robot social relationships.

My response to this significant shift in perspective will target and address three items: First, I want to connect-the-dots between the concept of "rhetorical agency" that is mobilized in the essay and recent innovations in the field of communication studies, specifically a new research paradigm called "Human Machine Communication" (HMC). Second, I will examine how the Confucian "role-based ethics" that is profiled in the text not only reproduces the decisive pivot that organizes the "social relational ethics" that Mark Coeckelbergh and I have developed and interjected into the field but also supplements those efforts by providing something that has been perceived to be absent or underdeveloped with the "relational turn." Third, I will end with a corrective or, more accurately stated, a rebuke, which is offered not in an effort to identify and call attention to fault but to assist the further cultivation and refinement of the essay's argument. All three comments, then, are designed to open lines of communication and to facilitate dialogue. They are, in other words, presented here in the spirit of what the essay characterizes as "Confucian friendship."

✉ David J. Gunkel
dgunkel@niu.edu

[1] Department of Communication, Northern Illinois University, Reavis Hall, 112, Dekalb, IL 60115, USA

## Communication and Artificial Intelligence

Unlike the majority of published work in the field of AI/robot ethics, which typically begins by inquiring about the characteristics or internal properties of the robot in order to determine whether it is or is not a legitimate moral agent, "Blame-Laden Moral Rebukes" focuses attention on the social circumstances in which the robot is situated and operates. What matters, therefore, is not what the robot *is* (an ontological issue), but on how it functions, especially in situations "where natural language capabilities may lead humans to intuitively ascribe social moral agency, a status that comes with unique persuasive powers." For all the attention that is paid to questions of moral agency in the literature, there is little or nothing concerning this kind of "rhetorical agency," which, as Cheryl Geisler (2004, 10) points out, is not beholden to the standard modernist concept of an autonomous (Cartesian) subject.

When considered from this other perspective—one that begins with and proceeds from the moral ecology of human–robot relationships—what really matters and what makes the difference is *communication*, specifically "the influence of robot communication strategies on the moral development of human teammates." In other words, what the robot really is turns out to be less important than the roles it comes to occupy as a socially interactive other in human–robot relationships. Unfortunately, the one discipline that would be well situated to investigate and develop this innovative insight—namely, the field of communication studies—has not adequately recognized or responded to this opportunity/challenge. But it should have.

Back in 1985—when personal computers were in their infancy and the Internet was little more than a National Science Foundation project for connecting academic research institutions—two communication scholars, Robert Cathcart and Gary Gumpert, sought to figure out and explain the significance of the computer for human social associations. In an essay titled "The Person–Computer Interaction," the two researchers distinguished between communicating *through* a computer and communicating *with* a computer. The former describes what is now widely recognized as Computer-Mediated Communication (CMC), which has been one of the dominant research paradigms in the field since the turn of the century. The latter, which had been largely neglected for decades, only recently began to receive attention under the moniker "Human–Machine Communication" (HMC).

Unlike efforts in Human–Robot Interaction (HRI), which concerns the design and operations of the control "interface" that is situated between human user and robotic instrument, HMC responds to the face—or the social facing—of the machine. As Andrea Guzman (2018, 3) explains in her agenda-setting collection of essays in HMC, "in human–machine communication, technology is conceptualized as more than a channel or medium: it enters into the role of a communicator." In occupying the role of a communicator, the machine is in a position to influence and persuade human teammates in ways that could have an important impact on their moral character, conduct, and development. The word "could" is

important here, and the "Blame-Laden Moral Rebukes" essay explicitly recognizes that "further empirical studies" will be necessary to test and verify whether the Confucian-inspired ethics that it espouses will actually help to develop this capacity in human teammates or not. HMC, with its focus on the machine as interlocutor and rhetorical agent is perfectly situated to respond to this need and to take responsibility for generating the data necessary to fill in the blanks and advance the debate.

## The Relational Turn

The Confucian role-based ethics that is presented and developed in the course of the essay provides another articulation of the moral innovation that I (and others) have called "the relational turn." But not just that. It also supplements those efforts by explicitly responding to what has been identified as something of a lacuna or open question in that research.

Mark Coeckelbergh and I (working independently on opposite sides of the Atlantic Ocean) published two books on the subject of AI/robot ethics that argued for a shift in the way that we decide questions of moral subjectivity—*Growing Moral Relations: Critique of Moral Status Ascription* (Coeckelbergh 2012) and *The Machine Question: Critical Perspectives on AI, Robots and Ethics* (Gunkel 2012). Within Western European philosophical traditions, the moral status of others is commonly decided and conferred based on the ontological properties of the entity in question. As Luciano Floridi (2013, 116) has explained, "what the entity is determines the degree of moral value it enjoys, if any." According to this standard procedure, the question concerning the status of others—whether they are someone who matters or something that does not—would need to be resolved by first identifying which property or properties would be necessary and sufficient for moral status, and then figuring out whether a particular entity possesses that property or not." The "relational turn" flips the script on this procedure; moral status is decided and conferred not on the basis of subjective or internal properties but according to objectively observable, extrinsic social relationships. "Moral consideration," as Mark Coeckelbergh (2010, 214) described it, "is no longer seen as being 'intrinsic' to the entity: instead it is seen as something that is 'extrinsic': it is attributed to entities within social relations and within a social context."

Both Coeckelbergh and I developed this alternative way of resolving questions of moral standing from within the Western European philosophical tradition—Coeckelbergh by following innovations in environmental ethics, specifically the work of J. Baird Callicott, and myself by calling upon Jewish philosophical traditions, especially "the ethics of otherness" that was the hallmark of Emmanuel Levinas, Jacques Derrida, and others (Gunkel 2018). Confucian ethics provides another way to approach this subject matter, proceeding from outside the Western European tradition and formulating a role-based ethic that is context dependent. For Confucian ethics (at least as it is developed and presented in the article), what matters is not what the robot is—its ontological properties—but "the roles the robot assumes and

the relationships the robot has with its human teammates in specific temporal and spatial contexts."

But it would, I think, be impetuous to conclude that Confucian role-based ethics is just another version or turn of what Anne Gerdes (2015, 274) has called the "relational turn" in ethics. Doing so risks instituting that kind of cultural appropriation that has been the dark underside of modern European thought—a violent reduction of the other to the same, as Levinas (1969) would have described it, or the domestication of the "exotic other" by way of what Edward Said (1979) identified with the term "orientalism." Furthermore, such reductionism misses an important opportunity and insight, specifically the way that Confucian role-based ethics can *supplement* the relational turn. "Supplement" understood in terms of the complex denotation that Jacques Derrida imparted to the word: "The supplement is an addition from the outside, but it can also be understood as supplying what is missing and in this way is already inscribed within that to which it is added" (Bernasconi 2014, 19).

One criticism of the "relational turn," a criticism that John Danaher (2019) has articulated rather well, is that this alteration in the way moral status is decided and ascribed does not necessarily provide (nor is it intended to provide) clear ethical guidance concerning the treatment of others. In other words, the relational turn seems to lack a normative dimension. Confucian ethics is able to respond to and supplement this perceived deficiency, contributing a "stronger emphasis on the psychological dimension of morality" and the resultant moral cultivation and refinement of the human participants. This role-based ethics, therefore, can help add on to and fill-out some of the perceived gaps in the relational turn, accounting for how the "inner psychological state" of the human teammates—specifically their perceptions of the robot's social roles and performances—necessitate specific ethical responses. This aspect has not been fully developed or appreciated in the current formulation of relational ethics, and the Confucian perspective opens the opportunity to supply a more complete picture of how things operate on the ground.

## Terminological Miscalculations

I conclude with a criticism that is less a "complaint" about the text and more a "rebuke" that is offered in order to assist its cultivation and refinement. The rebuke targets a crucial misunderstanding of technical terminology and concerns the following statement: "In long term interaction with their human teammates, or what computer scientists would call 'deep learning,' social robots might be able to develop 'moral knowledge' that can be transferable to other similar contexts." There are at least two problems here.

First, what is described in this sentence is specifically *not* what computer scientists call "deep learning." The term "deep learning" refers to a specific type of artificial neural network (ANN) where "deep" describes the depth (or number) of hidden layers in the network. Generally speaking, there are two methods for developing AI applications: symbolic reasoning or what is also called Good Old Fashioned AI (GOFAI) and neural network connectionist architectures that support machine learning (cf. Gunkel 2020). Unlike GOFAI programs, which need to code explicit

step-by-step instructions in executable statements, neural networks consist of a web of interconnected logic gates that are arranged in a hierarchy of different layers and that can be configured or tuned through training on data.

Whereas the programmers of a GOFAI algorithm need to anticipate and code for every conceivable situation the device may encounter, developers of ANNs only need to setup the network and select the training data and methodology. Currently, there are a number of different methods for achieving this with names like "supervised learning," "unsupervised learning," "reinforcement learning," etc. "Deep learning," by contrast, is not a method of machine learning. It describes the architecture or arrangement of neurons in the ANN where there are a large number of arrays of neurons in between the network's input and output layer. What the author of the essay describes, namely "long term interaction with human teammates," might be one kind of method for developing a form of "reinforcement learning," but it is certainly not "deep learning," at least not as far as this term is utilized by computer scientists.

Second, the latter part of the sentence is also troubling: "social robots might be able to develop 'moral knowledge' that can be transferable to other similar contexts." What a deep learning algorithm "develops" is not "knowledge" as we typically understand the word. What it develops is a set of weighted connections between the artificial neurons of the network that are, due to this particular configuration, able to transform input into a suitable output by operationalizing statistical patterns that are discoverable in the training data. Calling this procedure "knowledge" (even when set-off in scare quotes) is probably going too far and, what is worse, risks invalidating the important innovations that the essay introduces and describes. Stipulating that robots might develop or possess "moral knowledge" that can be transferable across contexts seems to reintroduce the Western philosophical obsession with generalizable moral principles that the Confucian role-base ethics calls into question and seeks to avoid. In other words, this statement risks reinstituting and falling back into the standard moral systems that the entire essay so successfully sought to question in the first place, thus undermining its own innovations and conclusions.

## References

Bernasconi, R. (2014). Supplement. In C. Colebrook (Ed.), *Jacques Derrida: Key concepts* (pp. 19–22). New York: Routledge.

Cathcart, R., & Gumpert, G. (1985). The person–computer interaction: A unique source. In B. D. Ruben (Ed.), *Information and behavior* (Vol. 1, pp. 113–124). New Brunswick: NJ: Transaction Books.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology, 12*(3), 209–221. https://doi.org/10.1007/s10676-010-9235-5.

Coeckelbergh, M. (2012). *Growing moral relations: Critique of moral status ascription*. New York: Palgrave MacMillan.

Danaher, J. (2019). Welcoming Robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-019-00119-x.

Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.

Geisler, C. (2004). How ought we to understand the concept of rhetorical agency? Reports from the ARS. *Rhetoric Society Quarterly, 34*(3), 9–17.

Gerdes, A. (2015). The issue of moral consideration in Robot Ethics. *ACM SIGCAS Computers and Society, 45*(3), 274–279. https://doi.org/10.1145/2874239.2874278.

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, Robots, and Ethics*. Cambridge: MIT Press.
Gunkel, D. J. (2018). *Robot rights*. Cambridge: MIT Press.
Gunkel, D. J. (2020). *An introduction to communication and artificial intelligence*. Cambridge: Polity Press.
Guzman, A. L. (2018). *Human-machine communication: Rethinking communication, technology and ourselves*. New York: Peter Lang.
Levinas, E. (1969) *Totality and infinity: An essay on exteriority.* Trans. Alphonso Lingis. Pittsburgh, PA: Duquesne University.
Said, E. (1979). *Orientalism*. New York: Vintage Books.