5-29-2019

Confucian Robot Ethics

Qin Zhu
Colorado School of Mines

Tom Williams
Colorado School of Mines

Ruchen Wen
Colorado School of Mines

Confucian robot ethics

Qin Zhu
Colorado School of Mines

Tom Williams
Colorado School of Mines

Ruchen Wen
Colorado School of Mines

Abstract

In the literature of artificial moral agents (AMAs), most work is influenced by either deontological or utilitarian frameworks. It has also been widely acknowledged that these Western "rule-based" ethical theories have encountered both philosophical and computing challenges. To tackle these challenges, this paper explores a non-Western, role-based, Confucian approach to robot ethics. In this paper, we start by providing a short introduction to some theoretical fundamentals of Confucian ethics. Then, we discuss some very preliminary ideas for constructing a Confucian approach to robot ethics. Lastly, we briefly share a couple of empirical studies our research group has recently conducted that aimed to incorporate insights from Confucian ethics into the design of morally competent robots. Inspired by Confucian ethics, this paper argues that to design morally competent robots is to create not only reliable and efficient human-robot interaction, but also a robot-mediated environment in which human teammates can grow their own virtues.

Keywords: *Confucian ethics, Applied ethics, Robot ethics, Artificial moral agents, Comparative studies*

In the literature of artificial moral agents (AMAs), most work is influenced by either deontological or utilitarian frameworks (Vallor, 2018). It has also been widely acknowledged that these Western "rule-based" ethical theories have encountered both philosophical and computing challenges. Most notably, these frameworks often struggle to "accommodate the constant flux, contextual variety, and increasingly opaque horizon of emerging technologies and their applications" (Vallor, 2018, p. 209). To tackle these challenges, this paper explores a non-Western, role-based, Confucian approach to robot ethics. In contrast to the Western philosophical approaches to robot ethics (or ethics in general) that focus on "defin[ing] what the good is" and worry about "how one can come to know the good," Chinese philosophers represented by Confucian scholars are more interested in the problem of "how to become good" (Ivahoe, 2000, p. ix). Inspired by Confucian ethics, this paper argues that to design morally competent robots is to create not only reliable and efficient human-robot interaction, but also a robot-mediated environment in which human teammates can grow their own virtues.

We argue that exploring a Confucian approach to robot ethics has critical values in a variety of ways. For instance, philosophers have recently argued that Confucian ethics can provide other alternative ways of thinking about ethical issues associated with robotics. As argued by Pak-Hang Wong, "if the philosophy of AI and robotics only comes from the West, that won't be enough, because it won't always apply to non-Western countries… And you miss opportunities to think in different ways about technology" (Cassauwers, 2019). In other words, Confucian ethics along with other non-Western ethical resources can enrich the moral imagination of roboticists and enhance their capabilities to define and engage ethical issues in designing robotics from culturally diverse perspectives.

Another crucial value for studying Confucian robot ethics is concerned with the powerful role that China and other cultures with the Confucian heritage (CHCs) such as Japan, Korea, and Singapore assume in the global market and the global robotics community. Understanding Confucian ethics is critical for understanding how policymakers, industrial entrepreneurs, scientists, and the public in CHCs view and accompany robotics.

In this paper, we start by providing a short introduction to some theoretical fundamentals of Confucian ethics. Then, we discuss some very preliminary ideas for constructing a Confucian approach to robot ethics. Lastly, we briefly share a couple of empirical studies our research group has recently conducted that aimed to incorporate insights from Confucian ethics into the design of morally competent robots.

Confucian ethics: Theoretical fundamentals

In recent decades, philosophers have employed various approaches to engaging Confucian ethics ranging from overtly historical or textual approaches to comparative approaches that put ideas from the classical period into conversation with contemporary Western ethical, social, and scientific theories (Mattice, 2019). Scholars have tried to understand Confucian ethics as a species of deontology, virtue ethics, or care ethics (Mattice, 2019). Until very recently, scholars have attempted to theorize Confucian ethics as a kind of role-based moral theory (Ames, 2011; Rosemont & Ames, 2016). The role-based approach to Confucian ethics is one of most recent efforts to reinterpret and rediscover the value of Confucian ethics. Confucian role ethics argues that we as humans were born into a web of social relationships. These social relationships have normative implications and they prescribe specific moral responsibilities for us in the communities we belong to. Ames (2011) argues that the term person itself is *relational* and *social* (e.g., it is hard to call a human "person" if she is the only human in the world). For instance, the father-daughter relationship does not only have descriptive values (e.g., describing an objective relationship between me and my daughter) but also has normative implications (e.g., what a healthy father-daughter relationship looks like, what expectations about being a good father are, what I should do to live my role as a father). Therefore, the Confucian ultimate goal of becoming a good person depends on to what extent we live well our social roles and practice well the moral responsibilities prescribed by these social roles.

Relationships, contexts, and social roles are crucial in Chinese philosophy especially Confucian ethics. The cultivation of the moral self including various virtues such as the principal virtue *ren* (仁, benevolence, goodness, or humaneness) takes place in the development of relationships (Lai, 2017). A person seldom grows her virtues at home by herself through her own individual actions and reflections. Instead, reliable virtues are required to be cultivated, tested, and enhanced in her *interactions* with others in specific contexts while living her communal roles. As a father, I learn virtues that define a good father not from reading books but from interactions with my daughter. The term father is coexistent with the term daughter. My daughter provides me with opportunities to develop the virtues that are required for being a good father. Cultivating virtues is thus a project that is engaged in concert with others.

Compared to most Western ethical approaches that focus on moral reasoning and justification, Confucian ethics places more emphasis on moral practice and practical wisdom. What is central to Confucian ethics is the moral development model that consists of three interrelated components: observation, reflection, and practice (Zhu, 2018). In other words, one must carefully observe how people in the society interact with others and reflect on whether and how their daily interactions with others are in accordance with *li* (礼, rituals or ritual propriety). The appropriate practice of rituals manifests virtues, whereas virtues underlie and guide the practice of these rituals. Then, one needs to incorporate her reflective learning experience into her own future interactions with others and test to what extent she has grasped the appropriate practice of rituals and their underlying moral virtues.

For Confucians, *li* or rituals are crucial for ethics practice and they are the social norms that are rooted in historical traditions and have been widely recognized as morally accepted norms in specific cultural contexts (Lai, 2003). Therefore, virtues (e.g., *ren*) and *li* are independent of each other (Lai, 2017; Shun, 1993). To demonstrate that one understands well a virtue such as *ren*, one needs to express it through appropriate ritual practice in her interactions with others. Conversely, her manner of interaction with others indicates her grasp of the virtue of *ren*. In this sense, Confucian criteria for evaluating one's moral development are *social* as they depend on to what extent people comprehend and appropriately practice rituals in their social interactions. In other words, someone who fails to or is unwilling to reflectively practice rituals misses opportunities for moral growth and thus is not a responsible person.

If the ultimate goal for Confucians is that one is *always* striving to become a good person through reflective learning in social interactions, then the question becomes whether everyone has the equal opportunity to achieve such goal. As a follower of Confucius, Mencius advocated for a kind of moral egalitarianism and believed that all human beings have the equal potential to become good if they devote themselves to intentional moral efforts (Munro, 1969, p. 15). To Confucians, what characterizes the personhood is not so much about one's innate and inalienable individual human rights as most Western political and ethical theories would emphasize. Instead, Confucians think that it is one's intentional efforts to strive for a good person that defines her personhood. Simply eating and sleeping without much reflective thinking will not make someone a *true* person (at least it is not a kind of person whose life is worth living).

Toward a Confucian robot ethics

A recent essay published in *OZY* by Tom Cassauwers (2019) aimed to reexamine ethical issues associated with robotics from two Eastern schools of thought: Confucianism and Buddhism. This essay includes interviews he conducted with multiple Eastern philosophers including two Confucian scholars Pak-Hang Wong and Heup Young Kim. An important theme in Confucian ethics of technology acknowledged by both Wong and Kim is that technology is never value neutral and it has crucial *instrumental* value that helps people acquire virtues such as *ren* and cultivate the moral self (Cassauwers, 2019). For instance, Wong mentioned an example in a recent redesign of Amazon's virtual assistant Alexa: designers have developed a new feature "politeness feature" for Alexa which will make Alexa only respond to people who speak to Alexa politely. Wong argues that such minor design could be made by a Confucian (Cassauwers, 2019).

Nevertheless, arguably, there remains a question whether these Amazon designers were *actually* inspired by Confucian ethical theories or more specifically Confucian moral psychology. Such question is worth exploring as it is relevant to the argument discussed earlier that relationships and contexts are important for Confucian ethics. The effectiveness of Alexa's politeness feature may be dependent on the specific role Alexa plays in a context and the relationship between Alexa and the human interlocutor. I as the father refuse to respond to my daughter's impolite request might have different moral impacts on her than if a stranger does the same reaction. Philosophers of technology such as Peter-Paul Verbeek would agree that design engineers do have the obligation to imagine the potential relationship that will be constructed between technology and its user and how such relationship affects the moral perception and behavior of the user (Verbeek, 2006).

What people are often overlooking is the *relational* nature embedded in the design of most robots especially social robots which are being integrated into our society. When robots are being designed, certain relationships are imagined, defined, and assigned to those robots and these relationships are often determined by the use contexts of robots and the specific roles robots are expected to assume in these contexts (e.g., healthcare robots). Dumouchel and Damiano (2017) recently argue that social robots such as Geminoid and Paro can only truly interact with other agents, and not with objects. Unlike humans, these robots have no relation to the world but to their human partners. These robots were mainly created for the interaction or relationship with human partners. It is the interaction or relationship between robots and their human partners that makes the *existence* of these robots. In this sense, we suggest that roboticists should not only leverage the traditional, dominant approaches to developing AMAs that focus on integrating rule-based morality, but also consider an alternative approach to designing morally competent robots based on the *role responsibilities* prescribed by the relationships robots have with human teammates in specific use contexts.

Philosopher JeeLoo Liu (2017) constructed three principles for Confucian robotic ethics that are based on the role moralities of robots:

[CR1] A robot must first and foremost fulfill its assigned role.

[CR2] A robot should not act in ways that would afflict the highest displeasure or the lowest preference onto other human beings, when other options are available.

[CR3] A robot must render assistance to other human beings in their pursuit of moral improvement, unless doing so would violate [CR1] and [CR2]. A robot must also refuse assistance to other human beings when their projects would bring out their evil qualities or produce immorality.

Liu's three Confucian robotic ethical principles well integrate major elements of Confucian ethics we introduced in the last section. A moral competent robot is expected to be able to fulfill its assigned social roles. Such fulfillment of social roles for this robot is not isolated from but in concert with other humans. For humans, self-cultivation is not an individual but a social project which the robot can contribute to. In other words, the interaction between the robot and its human teammate is indeed crucial for the human's pursuit of moral development. A socially integrated robot is expected to be sensitive to the norms shared within human communities and contribute to the evolution of these norms.

Confucius would probably argue that a social robot who is not capable of rendering assistance to humans in their pursuit of moral improvement is not worth being a companion for humans. Such judgment of the moral quality of the robot is comparable to Confucius's thesis that moral cultivation is essential to friendship. Interestingly, philosophers David Hall and Roger Ames (1994) argue that Confucian friendship is hierarchical despite that the hierarchy in friendship is different from that in other four Confucian relationships such as the father-son relationship. The hierarchy that exists in friendship recognizes that difference exists in the level of moral excellence between oneself and her friend (Lu, 2010). That is partly why Confucius said "do not accept as a friend one who is not your equal" (*Analects*, 1.8). Here, friendship as a relationship does have *instrumental* value, that is, a good or worthwhile friendship often provides opportunities and resources for the cultivation of the moral self. If we treat friendship as a paradigmatic case for the relationship between most social robots and their human teammates, then shouldn't we always strive to find robots who are capable of making us better people? Social robots may be distinguished in terms of their different capabilities of completing tasks. However, we argue that they should also be distinguished by their different capabilities of exerting positive impacts on the moral development of their human teammates.

The emerging literature in responsible innovation suggests that design is a "far richer process" as it realizes both functional requirements and moral values." Designers not only "can provide us with technical means but also can address the values of people and society and think about expressing them in material culture and technology" (van den Hoven, Vermaas, & van de Poel, 2015, p. 3). Therefore, roboticists should not simply consider robots as efficient means to help human users complete tasks. Moreover, for designing a social robot, roboticists need to consider other morally relevant issues such as:

- What social role is such robot expected to assume in its use context?

- What are characteristics or "traits" of this robot that defines it as a morally competent or "good" robot? How does the assigned role of the robot prescribe or specify these characteristics or "traits"?
- What kind of person is the human teammate becoming through her everyday interaction with the robot?

Social robots can be considered as companions with whom humans spend a lot of time. Nevertheless, does that mean a truly socially integrated robot has always to be polite and please its human teammate, even when the human teammate proposes morally questionable requests? Alistair M. C. Isaac and Will Bridewell (2017) point out that "standing norms" (baseline rules for effective human conversations such as "being polite or informative") are important for robots to be truly socially integrated and effectively communicate with humans. However, they also note that in meaningful conversations *ulterior motives* often are more fundamental to and thus supersede these standing norms. In other words, when designing strategies robots employ to respond to human requests, roboticist need to consider the ulterior motives (e.g., being a good companion) that are communicated through robot responses to human requests, in addition to the standing norms that are expected to be followed by robots and human teammates.

Therefore, is it okay for robots to blame morally questionable requests proposed by their human teammates as a person would do to another person (e.g., a friend)? From the Confucian perspective, one may argue that the way in which the robot responds to the human request is highly *contextual* and it depends on a variety of factors such as: under what context the human teammate asks the request, what role this robot plays, what relationship the robot and the human teammate has established, the level of ethical sensitivity the human teammate exhibits, and how much the robot "knows" the nature or personality of its human teammate. It is also worth noting that the relationship between the robot and its human teammate may also change as they interact with each other on the daily basis. For instance, if the relationship between a robot and its human teammate is reliable and trustworthy which is comparable to friendship in the Confucian sense, then it might be justifiable that the robot should be able to remonstrate with or blame its human teammate. Arguably, the role responsibility of the robot prescribed by its relationship with the human teammate encourages the robot to be responsible for the moral development of *the other one* who also contributes to such relationship. This argument is supported by Liu's third Confucian robotic ethical principle we mentioned earlier.

To some extent, despite the importance of role responsibility in Confucian ethics, Western philosophers hold different views on the connection between social roles and relationships and autonomous moral agency. To Western philosophers such as Dumouchel and Damiano (2017), the social roles and relationships assigned by roboticists to robots make robots less independent and thus have less moral agency which is fundamental for most Western political philosophical concepts such as liberty and autonomy. As discussed earlier, social robots are often *designed* to work in specific circumstances and serve certain purposes for humans. They are not independent and do not have or pursue their own goals. Dumouchel and Damiano (2017) argue that only robots with no explicit purpose may have autonomous moral agency comparable to

human personhood. Not having an explicit and predetermined purpose indicates that these robots are *free* to do anything they want.

In contrast, Asian philosophies pay less attention to the individualistic and liberal assumption of moral agency or personhood and instead they place more emphasis on the importance of social roles and relationship to personhood. Arguably, Asian philosophies such as Confucianism and Buddhism may provide possibilities "for nonhumans [such as robots] to reach the status of humans" (Cassauwers, 2019). As argued by Wong,

> In Confucianism, the state of reaching personhood is not a given. You need to achieve it. The person's attitude toward certain ethical virtues determines whether or not they reach the status of a human. That also means that we can attribute personhood to nonhuman things like robots when they play ethically relevant roles and duties as humans (Cassauwers, 2019).

Philosophical justifications for personhood in the West and the East may further lead to cultural differences in the public perceptions of robots. Spanish philosopher Jordi Vallverdú notices the cultural differences in human perceptions of robots between East and West: "Westerners are generally reluctant about the nature of robotics and AI, considering only humans as true beings, while Easterners more often consider devices as similar to humans" (Cassauwers, 2019).

Confucian ethics and designing morally competent robots

In this section, we are trying to provide some practical examples that demonstrate possible ways in which Confucian ethics can help to understand, inform, and shape the design of robots, and how this design process can help us refine our understanding of Confucian ethics.

One of the fundamental activities undertaken by robot designers is the identification, refinement, and application of *design patterns* (Alexander, et al., 1977; Borchers, 2000; Kahn, et al., 2008): abstract patterns of human interaction with the physical and social world that can be flexibly instantiated, nested, and combined. Kahn et al. (2008) list *claiming unfair treatment or wrongful harm* as one of the key design patterns for social robots, alongside common activities such as initial introductions, didactic communication, and recovering from mistakes, and describe how this key interaction pattern of *protest* can be instantiated from both deontological and consequentialist perspectives in order to assert a robot's moral standing.

This design pattern is itself just one example of the broader design pattern of *identifying* unfair treatment or wrongful harm, i.e., protesting an action not necessarily on the basis of unfairness or harm towards oneself, but more generally on the basis of some identified unfairness or harm. In our own work, we have examined the tradeoffs between different instantiations of this design pattern. Specifically, we have looked at different *Speech Act-theoretic* (Searl, 1969) that robots might reject commands and requests when they are identified as harmful, examining the differential effects of

phrasing rejections as *questions*, *statements*, or *rebukes* (Jackson, Wen, & Williams, 2019).

But critically, as identified by Kahn et al., this design pattern can also be instantiated according to different ethical frameworks. Accordingly, this design pattern presents an excellent testing ground for applying and evaluating the effectiveness of different ethical frameworks. In recent work, we have accordingly begun to examine how humans perceive robot rejection of inappropriate commands when those rejections vary not only according to Speech Act theoretic phrasing, but also according to underlying ethical framework (Wen, Jackson, Williams, & Zhu, 2019). For example, if a robot serving as an instructor is asked to perform an action that constitutes or facilitates cheating, it may issue a question-phrased rejection in multiple ways: asking "Wouldn't that be cheating?" draws direct attention to the norm that would be violated if the directive were accepted, whereas "Would a good instructor do that?" instead draws attention to the robot's role, only directly highlighting the prohibitive norm. While our preliminary results suggest that these role-based norm violation responses lead humans to perceive robots as better fulfilling their professed roles, we have not yet found any evidence to suggest that role-based responses lead to any other effects that we might expect, such as increased mindfulness and self reflection. In our current work, we are designing experiments that more fully address cultural, contextual, and temporal considerations that may have prevented us from observing these other hypothesized effects. If our experiments elicit our hypothesized effects, this will serve as a strong argument in favor of robotic moral language generation grounded in role-based frameworks such as Confucian ethics.

Robot designers interested in enabling moral language generation grounded in Confucian ethics must make design decisions that articulate different positions and priorities within the Confucian perspective. First, Confucian ethics is a role-oriented paradigm, and thus designers must decide what roles are appropriate for robots to play within human society. In Confucian classics, five cardinal relationships (*wulun*, 五伦) are delineated: ruler-minister, father-son, husband-wife, older-younger, and friend-friend. For human-robot relationships, designers must articulate an equivalent set of cardinal human-robot relationships, e.g. supervisor-subordinate, owner-ownee, adept-novice, teammate-teammate, and friend-friend. Critically, a designer's choice of represented relationships may impact not only the contents of robots' norm violation responses, but the decisions as to whether those responses are generated in the first place. For example, from the point of view of Confucian ethics, if the *friend-friend* relationship holds between agents A and B, then A has the role ethics of remonstrating with B when A observes B committing or proposing wrongdoing: this remonstration is a requirement for A to be a good friend of B. Thus if designers choose to include *friend-friend* among the cardinal human-robot relationships they choose to represent, then this may affect the frequency with which those robots should choose to respond to proposed norm violations, and may impact the Speech Act-theoretic phrasing that robots should use in such circumstances (i.e., blame-laden moral *rebukes* may be necessary for proper remonstration when the friend-friend relationship holds between a robot and its observed norm violator).

Second, Confucian ethics espouses multiple competing objectives that may conflict with each other. For example, Confucian ethics emphasizes both self reflection

and emotional display of role commitment. When phrasing a rejection from a role-based perspective, it is possible to differentially encourage these different outcomes: by using a role-based interrogative (e.g., "Would you be a good friend if you did that?"), the robot may be able to encourage more self reflection; while when using a role-based rebuke ("You are a bad friend for asking me to do that!") the robot may instead demonstrate more emotional commitment to its role (and provoke a greater emotional response). The process of designing a robot that must respond to unacceptable commands thus forces us to think critically not only about what constitutes an ethical response, but also about how different aspects of an ethical framework may conflict with each other or be chosen between. Moreover, this provides us an opportunity to interrogate those ethical commitments. Ultimately, we must ask ourselves whether Confucian principles such as encouragement of self-reflection and emotional demonstrations of role adherence are the end goals we strive to achieve, or whether we are only seeking to achieve these goals because we believe they will lead others to take more role-fulfilling actions in the future. If our goal is the former, we can use the design, implementation, and evaluation of our computational models to identify whether those policies actually lead robots' interactants to achieve those principles; if our goal is the latter, we can further examine whether achievement of those intermediate principles actually correlates with achievement of our overarching societal goals. Such examination may allow the ethically-informed design process to feed back and inform the ethical framework itself, by quantifying the relative merits of the intermediate principles espoused by that framework.

References

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., & Shlomo, A. (1977). *A pattern language: Towns, buildings, construction.* New York, NY: Oxford University Press.

Ames, R. T. (2011). *Confucian role ethics: A vocabulary.* Honolulu, HI: The University of Hawai'i Press.

Borchers, J. O. (2000). A pattern approach to interaction design. *The 4th Conference on Designing Interactive Systems: Processes, Pactices, Methods, and Techniques* (pp. 369-378). New York: ACM Press.

Cassauwers, T. (2019, March 28). *How Confucian could put fears about artificial intelligence to bed.* Retrieved April 26, 2019, from OZY: https://www.ozy.com/immodest-proposal/how-confucianism-could-put-fears-about-artificial-intelligence-to-bed/93206

Dumouchel, P., & Damiano, L. (2017). *Living with robots.* Cambridge, MA: Havard University Press.

Hall, D., & Ames, R. T. (1994). Confucian friendship: The road to religiousness. In L. S. Rouner (Ed.), *The changing face of friendship* (pp. 77-94). Notre Dame, IN: University of Notre Dame Press.

Isaac, A. M., & Bridewell, W. (2017). White lies on silver tongues: Why robots need to deceive (and how). In P. Lin, R. Jenkins, & K. Abney (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 157-172). New York, NY: Oxford University Press.

Ivahoe, P. J. (2000). *Confucian moral self cultivation* (2nd ed.). Indianapolis, IN: Hackett Publishing.

Jackson, R. B., Wen, R., & Williams, T. (2019). Tact in noncompliance: The need for pragmatically apt responses to unethical commands. *The 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 499-505). Honolulu: ACM Press.

Kahn, P. H., Freier, N. G., Kanda, T., Ishiguro, H., Ruckert, J. H., Severson, R. L., & Kane, S. K. (2008). Design patterns for sociality in human-robot interaction. *The 3rd ACM/IEEE International Conference on Human Robot Interaction* (pp. 97-104). Amsterdam: ACM Press.

Lai, K. (2003). Confucian moral cultivation: Some parallels with musical training. In K.-c. Chong, S.-h. Tan, & C. L. Ten (Eds.), *The moral circle and the self: Chinese and Western approaches* (pp. 107-139). Chicago, IL: Open Court.

Lai, K. (2017). *An introduction to Chinese philosophy* (2nd ed.). Cambridge, UK: Cambridge University Press.

Liu, J. (2017). Confucian robotic ethics. *International Conference on the Relevance of the Classics under the Conditions of Modernity: Humanity and Science.* Hong Kong: Hong Kong Polytechnic University.

Lu, X. (2010). Rethinking Confucian friendship. *Asian Philosophy, 20*(3), 225-245.

Mattice, S. (2019). Confucian role ethics: Issues of naming, translation, and interpretation. In A. McLeod (Ed.), *The Bloomsbury research handbook of early Chinese ethics and political philosophy* (pp. 25-44). London, UK: Bloomsbury Academic.

Munro, D. (1969). *The concept of man in early China.* Stanford, CA: Stanford University Press.

Rosemont, H., & Ames, R. T. (2016). *Confucian role ethics: A moral vision for the 21st century.* Göttingen, Germany: Vandenhoeck & Ruprecht.

Searl, J. (1969). *Speech acts: An essay in the philosophy of language.* New York: Cambridge University Press.

Shun, K.-l. (1993). Jen and li in the Analects. *Philosophy East and West, 43*(3), 457-479.

Vallor, S. (2018). *Technology and the virtues: A philosophical guide to a future worth wanting.* New York, NY: Oxford University Press.

van den Hoven, J., Vermaas, P. E., & van de Poel, I. (2015). Design for values: An introduction. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of ethics, values, and technological design* (pp. 1-8). Dordrecht, Netherlands: Springer.

Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation. *Science, Technology, & Human Values, 31*(3), 361-380.

Wen, R., Jackson, R. B., Williams, T., & Zhu, Q. (2019). Towards a role ethics approach to command rejection. *The 1st HRI Workshop on the Dark Side of Human-Robot Interaction: Ethical Considerations and Community Guidelines for the Field of HRI* (pp. 1-4). Daego, Korea: ACM Press.

Zhu, Q. (2018). Confucian ethics, ethical leadership, and engineering ethics education. *International Journal of Ethics Education, 3*(2), 169-179.