

Role-based Morality, Ethical Pluralism, and Morally Capable Robots¹

Qin Zhu², Tom Williams³, and Ruchen Wen⁴

Dominant approaches to designing morally capable robots have been mainly based on rule-based ethical frameworks such as deontology and consequentialism. These approaches have encountered both philosophical and computational limitations. They often struggle to accommodate remarkably diverse, unstable, and complex contexts of human-robot interaction. Roboticists and philosophers have recently been exploring underrepresented ethical traditions such as virtuous, role-based, and relational ethical frameworks for designing morally capable robots. This paper employs the lens of ethical pluralism to examine the notion of role-based morality in the global context and discuss how such cross-cultural analysis of role ethics can inform the design of morally competent robots. In doing so, it first provides a concise introduction to ethical pluralism and how it has been employed as a method to interpret issues in computer and information ethics. Second, it reviews specific schools of thought in Western ethics that derive morality from role-based obligations. Third, it presents a more recent effort in Confucianism to reconceptualize Confucian ethics as a role-based ethic. This paper then compares the shared norms and irreducible differences between Western and Eastern approaches to role ethics. Finally, it discusses how such examination of pluralist views of role ethics across cultures can be conducive to the design of morally capable robots sensitive to diverse value systems in the global context.

Keywords: role ethics, ethical pluralism, morally capable robots, human-robot interaction, comparative ethics

1. Introduction

Dominant approaches to designing morally capable robots have been mainly based on rule-based ethical frameworks such as deontology and consequentialism. The moral quality of a robotic action depends solely on its consistency with ethical principles (e.g., categorical imperatives, the principle of utility). These rule-based ethical frameworks have both philosophical and computational limitations. They often struggle to accommodate remarkably diverse, unstable, and complex contexts of human-robot interaction (Vallor, 2016). Therefore, roboticists and philosophers have been exploring underrepresented ethical traditions such as virtuous, role-based, and relational ethical frameworks for designing morally capable robots (Coeckelbergh, 2010). Our research has drawn on intellectual resources from Confucian ethics to examine the impacts of role-based moral communication on the self-reflection and moral development of human teammates (Zhu, Williams, Jackson, & Wen, 2020).

¹ This work was funded in part by National Science Foundation grant IIS-1909847.

² Corresponding Author, Assistant Professor, Department of Humanities, Arts & Social Sciences, Colorado School of Mines, Golden, Colorado, USA. E-mail: qzhu@mines.edu

³ Assistant Professor, Department of Computer Science, Colorado School of Mines, Golden, Colorado, USA.

⁴ PhD Student, Department of Computer Science, Colorado School of Mines, Golden, Colorado, USA.

Nevertheless, to build up a robust role-based ethical framework complementary to existing rule-based frameworks, there are certain questions worth further investigation, especially if such a role-based framework is to be employed to design robots that can effectively perform tasks in a global context. For instance, does role-based morality solely belong to the Confucian tradition? Are there Western ethical traditions that care about role-based morality? If so, are these Western approaches compatible (or not) with Confucian role ethics? Will robots designed within the Confucian tradition function as expected (or not) in the Western context? To answer these questions, this paper critically reexamines the notion of role-based morality from the perspective of ethical pluralism. As a philosophical method, ethical pluralism acknowledges both resonances and radical differences between Western and Confucian views on role-based morality. Differences in the two traditions are sources of inspiration for gaining a deeper and more culturally sensitive understanding of role ethics. In contrast to ethical relativism, ethical pluralism emphasizes positive engagement between the two traditions and allows one tradition to enhance and elaborate on the characteristics of the other (Ess, 2006).

More specifically, this paper first provides a concise introduction to ethical pluralism and how it has been employed as a method to interpret issues in computer and information ethics. Second, it reviews specific schools of thought in Western ethics that derive morality from role-based obligations. Third, it presents a more recent effort in Confucianism to reconceptualize Confucian ethics as a role-based ethic. It then compares the shared norms and irreducible differences between Western and Eastern approaches to role ethics. Finally, it discusses how such examination of pluralist views of role ethics across cultures can be conducive to the design of morally capable robots sensitive to diverse value systems in the global context.

2. Ethical Pluralism and Intercultural Information Ethics

As a philosophical method, ethical pluralism is the view that “moral values, norms, ideals, duties, and virtues are irreducibly diverse” (Weinstock, 1998, Article summary section). In an increasingly globalized environment, it is critical that we consider whether and how ethical standards established can be responsive to diverse cultures. Neither ethical relativism nor ethical dogmatism would work for building ethical standards in the global context. After reviewing different approaches to ethical pluralism, Ess advocates that the *pros hen* (“towards one”) pluralism (interpretative or complementary pluralism) advocated by Plato and Aristotle can be helpful for “bridging the deep differences between Eastern and Western norms, values, and traditions” (Ess, 2006, p. 217). According to the *pros hen* pluralism, there can be multiple interpretations of one single idea and these different interpretations are “irreducibly different from one another but are connected and coherent with one another (not simply compatible) by way of their shared point of origin and reference” (Ess, 2006, p. 218). In contrast to ethical relativism, ethical pluralism emphasizes positive engagement between the two traditions and allows one tradition to enhance and elaborate on the characteristics of the other (Ess, 2006).

Ess (2006) is a major scholar who introduces the concept of ethical pluralism to global information ethics by comparing ethical pluralism in the West and in the East. To construct a global approach to information ethics, Ess analyzes the resonances and irreducible differences in concepts such as human nature, community, and privacy between Eastern and Western approaches to information ethics. Ess (2006) discovers both Eastern and Western approaches to information ethics share some similar views on human nature and community. For instance, both Aristotle and Confucius consider humans as relational beings who interact with each other in a

larger community. They both emphasize that the social order or harmony in the community starts with the family.

However, there are irreducible differences between Eastern and Western approaches to information ethics. Compared to a fundamental philosophy in contemporary Western societies – namely, liberalism that highlights the autonomous and atomistic self – a major Eastern philosophical tradition, Confucianism, values more a collective understanding of the human being and the self (Ames, 2011; Rosemont, 2015). With respect to privacy, for example, liberal democratic traditions in the West see both intrinsic value (e.g., individual privacy is intrinsically good and does not need any further justification) and instrumental value (e.g., privacy is instrumental for the development of personal autonomy). In contrast, the idea of individual privacy in Confucianism is mainly connected to some negative connotations (e.g., individual privacy means shameful secrets that are against communal interests and people hide from being known by others) (Ess, 2006).

Applying ethical pluralism to global information ethics raises key ideas, such as the shared value of privacy across cultures, despite diverse approaches to interpreting these ideas in different cultural and historical contexts. Holding an interpretative attitude toward ethical pluralism can be conducive for: (1) understanding diverse approaches to ideas shared by different cultures; (2) examining how these different approaches can benefit from each other; and (3) developing a more global account of values responsive to an increasingly globalized world.

In this paper, we adopt the lens of ethical pluralism to analyze both resonances and radical differences between contemporary Western and Confucian views on role-based morality.⁵ We further argue that resonances and differences in the two traditions are sources of inspiration for gaining a deeper and more culturally sensitive understanding of role ethics. Such an intercultural exercise that compares and engages these two intellectual traditions of role ethics is helpful for designing morally capable robots sensitive to diverse cultural values in the global context.

3. Role Ethics in the West

Role ethics is a recent effort in philosophy to question the fundamental idea of “autonomous individualism” in Western ethics. Roles have received very limited attention from analytic moral philosophers (Dare & Swanton, 2020). From a role ethics perspective, ethics should be theorized and practiced “from the standpoint of humans as teachers, parents, doctors, friends, and the like” rather than “humans as humans” (Dare & Swanton, 2020, p. 1).

Johnson (2014) analyzes Greek Stoic philosopher Epictetus’s (circa 50 to 130 A.D.) role ethics. Epictetus distinguishes two kinds of human roles: (1) the cosmic role: our universal, human role that requires us to admire God (e.g., Nature) and preserve God’s gift to us (e.g., our volition); and (2) local roles: the multiple roles we appropriately fulfill in the everyday life such as friend, teacher, sibling, guest, counselor, employee, and citizen (Johnson, 2014). Epictetus’s universal role focuses more on the concrete obligations of a good rational being and such a role calls for virtues such as rationality, trustworthiness, cosmopolitanism, sociality, piety, and the elimination

⁵ A major reason why the discussion of role-based morality places a strong emphasis on contemporary Western and Confucian approaches is that the existing literature on the topic of role ethics is extensively focused on the two traditions. Nevertheless, it is also worth noting that there is an increasingly interest among Western scholars in studying the role-based morality existing in ancient Greek philosophies such as Aristotelianism and Stoicism (Johnson, 2014).

of the passions (Johnson, 2014). Johnson (2014) also conceptualizes four criteria used by Epictetus to specify our local roles:

- particular nature: our (physical) abilities can limit and prescribe our specific roles (e.g., strong legs suggest an athletic career);
- social relations: both our natural relations (e.g., parent, sibling) and acquired relations (e.g., spouse, neighbor) define our specific roles;
- chosen roles: we choose some particular roles when our capacities allow us to pursue different possible careers;
- divine signs: divination can also indicate our roles.

Epictetus does emphasize that every role has meaningful obligations and represents a relation with others in a particular community. The roles we assume determine what is good and evil. Logic and reason are of critical importance in Epictetus's ethical theory (Johnson, 2014). It is through logic we can specify a particular role we assume and infer appropriate actions from such a role. It is because of reason that we formulate a reasonable defense for the role we assume. For instance, I can infer what I need to do from my role as an engineer. As an engineer, I have the moral obligation to promote public welfare through the services and goods I deliver. The code of ethics of the National Society of Professional Engineers (NSPE) in the United States stipulates that "engineers shall hold paramount the safety, health, and welfare of the public." As I am an engineer, it is reasonable for me to defend my role in providing services and goods to the public and promoting their welfare. Epictetus values the cosmic role over more particular local roles as "it is our relation to nature that makes possible our civic communities" (Johnson, 2014, p. 87).

What Epictetus calls local roles are the major themes often discussed in the Western literature in role ethics, although scholars have used different approaches to further analyze these local roles. In the literature, the term "role" mostly refers to Epictetus's local roles (rather than his conception of humans' universal role).

Philosopher George Herbert Mead (1863-1931) used the term "roles" in two senses: (1) an individual views the relationship between two or more other people within a context (e.g., a child takes her parent's role and mimics her parent's characteristic actions); and (2) a social position that any number of people might take up (e.g., a child plays at being a parent) (Pettigrove, 2020). Pettigrove (2020) calls Mead's second set of roles more "generic roles" which are the more usual meaning given to the term "roles" in social theory, compared to Mead's first set of roles which refer to our particular points of view. The generic roles are roles that are able to be taken by more people, whereas Mead's first set of roles refer to those that are observed by a particular, third person in a particular context.

Pettigrove (2020) further analyzes the more "generic roles" and emphasizes that these roles come along with normative expectations. Some of these generic roles are normatively thin and others are normatively thick (Pettigrove, 2020). Normatively thin roles are roles that only indicate what role-occupants characteristically do (e.g., the role of felon simply describes a person who is convicted of a crime but does not prescribe that such person should do bad things all the time). In contrast, normatively thick roles are those that not only describe what role-occupants characteristically do but also prescribe what they should do (e.g., the role of engineer does not

only describe the everyday practice of the engineer but also prescribe what the engineer should do).

When we apply normatively thick roles to others, we are also anticipating and evaluating what they will be doing. An engineer is expected to use her engineering knowledge to deliver services and goods to the public. If she fails to live up to her role that promotes the safety, health, and welfare of the public as required in engineering codes of ethics, she will be morally blamed. Our knowledge of the norms associated with these normatively thick roles can often mediate our actions and perceptions (Dare, 2020). Dare uses one example to demonstrate the mediating effect of roles. In his example of taking his child to a hospital's emergency department, he does not know anything about the person coming into the emergency room to treat his child. However, what he does know is the norms of the role as an emergency doctor. His knowledge of these "role-norms" will be likely to create expectations and predictions about the emergency doctor which he has never met. There are cases in which our knowledge of others and their social roles may not be very useful (e.g., sometimes we feel upset that a friend has done something disappointing which we did not expect) (Dare, 2020). Furthermore, roles are also attached with enablements and constraints (Pettigrove, 2020). A particular social role requires someone to accumulate some experience that enables her to be able to do something (e.g., the role of parent enables her to take care of her children). Meanwhile, the social role sometimes may also constrain a role-occupant from performing certain actions (e.g., the role of engineer constrains the engineer from performing services in fields that are not the areas of her competence).

Similar to Pettigrove, Korsgaard (2009) points out that we as rational beings constitute practical identities by occupying various social roles that are associated with certain moral duties. In discussing our social obligations, Korsgaard mainly distinguishes two kinds of roles: voluntarily adopted roles (e.g., engineer) and the roles connected with factually grounded identities (e.g., child) (Korsgaard, 2009). Like Epictetus, Korsgaard (2009) also briefly mentions more fundamental roles (similar to Epictetus's cosmic, universal role) enabled by our identity as rational beings. Influenced by Kantian ethics, concepts such as duties, reason, and categorical imperatives are central to Korsgaard's role ethics. As rational beings, we have the freedom to adopt and renounce our social roles (Korsgaard, 2009). We are free to renounce roles (except "non-contingent roles" made possible by being a rational human being) when these roles cause a conflict among our duties. An important way to identify the roles to renounce that cause duty conflicts is to appeal to Kantian deontology (Korsgaard, 2009). We as rational beings only adopt roles that are aligned with categorical imperatives and will not generate logical contradictions.

There are at least a couple of challenges with Korsgaard's approach. First, Korsgaard's approach seems to be more interested in the ethical theories underlying roles rather than roles *per se*. Dare calls such approach the derivative approach that "sees [roles] as part of the 'superstructure' of morality, always dependent, for both their content and their normative force, upon the foundational values of competing moral theories" (Dare, 2020, p. 31). In Korsgaard's case, Kantian deontological ethics is the ethical theory used to evaluate the moral significance of roles and role-norms are derived from the deontological theory. Second, Korsgaard's voluntarily adopted roles may not always be fully voluntary. Dare argues that some institutional roles such as those existing in law, medicine, or engineering may be voluntarily chosen by people but these roles are often created in a "process of authoritative institutional design that has attached a set of role-norms to [various] roles" (Dare, 2020, p. 35). Therefore, these norms are socially constructed and are the result of community consensus-making procedures. In contrast,

Korsgaard's factually grounded roles are "created in the absence of authoritative designers" (Dare, 2020, p. 35) such as the father's role.

If institutional roles are often created by deliberate institutional design and authoritative designers have attached norms to these roles, how are non-institutional, more "natural" and "social" roles (e.g., father, friend) often created? Where do the norms attached to these non-institutional roles come from? Dare points out that the norms associated with non-institutional roles are generated by "the complex social practices around roles within communities" and "the attitudes of community members toward them" (Dare, 2020, p. 37). Such an argument has strong cultural implications. Despite that the term father exists in many different cultures, the role ethics of father in a particular cultural context is determined by the social practices around the father's role in that community and how other members of the community perceive and react to such a role.

So far, it seems that most Western scholars have acknowledged the normative force of roles. Some of them such as Dare would further advocate that roles should be seen as more foundational to ethics, as moral obligations of role-occupants are derived from the roles they assume. However, Swanton rejects such a review and argues that virtues instead of roles should be more fundamental to ethics. "Role" does not mean "good role" and "we need a view about what makes for good roles and genuine role-obligations" (Swanton, 2020, p. 46).

In other words, a challenging question for scholars who advocate the foundational place of roles: how do we know a role is a good role for people to exercise and for others to imitate? How do we evaluate the moral quality of the "complex social practices" and "the attitudes of community members toward these practices" that generate role-norms as argued by Dare (2020)? There needs to be more foundational moral guidance. A role may be practiced and approved by most members in a community yet not be a good role.

4. Role Ethics in the Confucian Tradition

In Confucianism, role ethics "denotes a constellation of views ... that promote a relational conception of persons and employs this to emphasize how a person's roles and relationships are the source of ... ethical obligations and growth" (Ramsey, 2016, p. 235). Leading Confucian scholars Ames (2011) and Rosemont (2015) present role ethics as an alternative to rule following theories (e.g., deontology, consequentialism) and as a distinctive type of ethics.

In Confucian role ethics, our moral actions in different situations are shaped by the specific roles we take in these situations. We as humans all assume various roles which are determined by the relationships we have with others. These different social relationships and roles affect the ways we interact with others. For instance, the tone we use to speak to parents is different from the one we use to communicate with strangers. The nature of a particular role relationship often evokes feelings and expectations characteristic of that relationship (Ames, 2011). Roles do not simply describe our social relationships with others but also provide normative expectations about these relationships. In the first place, the roles Ames and Rosemont have in mind are family-based roles such as son, daughter, mother, older, sibling, and grandfather. Other social roles discussed in classic Confucianism such as ruler, subject, husband, wife, minister, and friend are also of interest to Ames and Rosemont (Angle, 2018). The discussion of role-based morality can be further extended from family roles to social or professional roles such as engineer, doctor, teacher, and nurse (Zhu, 2018).

Differentiation and fulfilment of social roles is critical for a harmonious society. In Confucian ethics, there are five cardinal role-based relationships and they are those between parents and children, husband and wife, older and younger siblings, rulers and ministers, and friends (Cottine, 2020). The five relationships belong to three social spheres: (1) family sphere: the parent-child relationship, the husband-wife relationship, and the relationship between siblings; (2) intermediary sphere: friendship; and (3) the social/political sphere: the ruler and minister relationship. Cottine (2020) argues that family relations are foundational for individual moral development and state governance. Being a filial child provides a paradigmatic case for being a loyal minister.

Through living and reflecting on these social roles, we cultivate virtues that are necessitated by the practice of these social roles. To live and reflect on the role as a medical doctor, one gets to cultivate virtues (e.g., benevolence) that are required by being an ideal medical doctor. Nevertheless, such process of cultivating virtues cannot be solely completed by the doctor herself. It needs to be done by both the doctor and the patients. Therefore, Confucian role ethics advocates a kind of “relational virtuosity”: becoming benevolent is something we either do together, or not at all (Ames, 2011). Moral conduct thus refers to behavior that “conduces to growth in the roles and relations we live together with others, and immoral conduct is the opposite” (Ames, 2021). Confucian role ethics acknowledges the value of social roles in making an agent the person she is (Nuyen, 2007). It is one’s intentional efforts to actively live her social roles that defines her personhood.

Therefore, Confucian role ethics defines humans as “the sum of the roles we live in consonance with our fellows” (Ames & Rosemont, 2011, p. 20). Confucian role ethics appeals to the actual life experience we are living with others both cognitively and affectively (Ames & Rosemont, 2011). A critical way of becoming virtuous persons is to observe how others practice *li* (rituals, 禮) that are required by their social roles. Practicing rituals appropriately can be conducive to the reinforcement of human relationships and associated communal roles. Ritual practices require us to both physically and emotionally engaged (Hagen, 2010). Emotions and feelings are critical for demonstrating our commitment to the practice of rituals and the fulfillment of our role-based moral obligations. A truly caring nurse can never be one who only knows how to follow rules. She develops her benevolence by feeling what her patients are suffering. Arguably, her emotional engagement with patients’ experience encourages her to develop qualities that define a truly caring nurse. Thus, one possible way of evaluating whether we fulfill our communal roles well is to examine whether we have any emotional investment in these roles.

Bell distinguishes two approaches to the relation between Confucian role ethics and morality: “the strong claim that the (constitutive) roles we occupy determine the content of our moral obligations” and “the weak claim that our (constitutive) roles set constraints upon what we ought to do” (Bell, 2018, p. 206). Unlike Ames who strongly advocates the strong claim, Bell argues that the weak claim is more persuasive. Bell further points out that Confucian role ethics needs to be constrained by moral standards external to the roles if it is to provide morally informed practical guidance, as awareness of one’s role *per se* may not provide clear guidance on moral actions (Bell, 2018). Bell articulates certain reasons for why moral standards external to social roles are more fundamental. First, the moral obligations associated with particular social roles can change over time. Second, there will be cases in which our different roles can lead to conflicting moral obligations (Bell, 2018). Therefore, it is worthwhile to fulfill the moral

obligations prescribed by our social roles insofar as doing so will not violate any fundamental principles for the harmony of the society (e.g., human rights). Social roles may not always provide clear and detailed guidance for moral actions particularly in situations that involve conflicting obligations, however, social roles can be helpful for telling us what we cannot do (Bell, 2018).

Nevertheless, defenders of Ames and Rosemont's approach such as Ni argue that Ames and Rosemont have at least partially addressed Bell's concern. Ames and Rosemont emphasize the "highly particularistic" approach to understanding the moral implications of social roles (Rosemont, 2015). They reject any attempt to consider Confucian role ethics as a theory that provides preset moral standards for actions (Rosemont, 2015; Rosemont & Ames, 2016). Ni (2018) points out that the social roles themselves do not fully predetermine what we ought to do. For Confucian role ethics, it would be dangerous to assume a metaphysical position that humans are metaphysically relational beings (Ni, 2018). If we accept the metaphysical position, we need to predetermine the goodness of each social role. Presuming a single preset right way of assuming these different social roles can be a challenging task. According to Ni,

[Confucian role ethics] is not a theory that determines rights and wrongs according to preset standards of abstract, generalized roles (like the "three obedience" doctrine), but rather determines them through particular interactions within the context of the roles that relevant persons live. (Ni, 2018, p. 194)

Higgins (2018) feels worried about Ames's extremely idealistic vision of family, family roles, and other social roles extended from family roles. She worries that Confucian role ethics might reinforce social roles that are oppressive. Strong demands for conformity to social roles can potentially lead to "oppression of sexual minorities and others who perform their roles atypically" (Higgins, 2018, p. 218). It is worth noting that Higgins's concern can also be found in Western scholars' criticisms of role ethics (Swanton, 2020). The next section will systematically compare the Western and the Confucian approach to role ethics.

5. Role Ethics in the Intercultural Context

This section mainly compares the shared norms and irreducible differences between Western and Confucian approaches to role ethics. The last two sections show that both Western ethics and Confucianism see role ethics as a possible effort to revitalize a philosophy that deserve more attention from scholars in mainstream philosophical traditions. As such, the concept of roles has some shared value and is viewed similarly between the two traditions.

Both Western ethics and Confucianism notice that the diverse range of roles we live in include both more natural, personal, social roles (e.g., father, child, sibling, spouse, friend) and more institutional and professional roles (e.g., minister or public servant, doctor, engineer). Both traditions have acknowledged not only the descriptive aspect but also the normative aspect of role. Western and Confucian scholars both agree that a role does not only describe a relationship but also provides normative content for such role. Roles do create normative expectations for us.

Scholars from both Western and Confucian traditions have challenged whether roles are fundamental to ethics and to what extent roles can provide clear normative guidance for actions. From the Western perspective, Korsgaard (2009) argues that Kantian deontological ideas such as categorical imperatives are more fundamental than roles and roles causing conflicts among our

duties should be renounced. Korsgaard's view is close to what Dare would call the derivative view that derives role-norms from ethical theories such as deontology. Swanton (2020) distinguishes between role vs. good role and argues that roles are just facts and roles themselves do not provide clear ethical guidance for ethical conduct. Instead, she argues that virtues are more fundamental than roles to ethics. From the Confucian perspective, Bell (2020) suggests that moral standards external to the roles are needed as our awareness of roles *per se* does not provide clear, practical guidance as (1) the meaning of our roles can always change; and (2) there can be role conflicts. Higgins (2018) feels concerned about the idealistic vision of family in the Confucian tradition and worries that unexamined social roles may lead to oppression of marginalized groups.

Nevertheless, it will not be difficult to notice that there are irreducible differences between Western ethics and Confucianism in understanding the moral significance of role. The two traditions differ in terms of what constitutes the essence of role. Western ethics takes a more metaphysical stance on the essence of role and acknowledges that the roles humans assume are ready to be reasoned about, critiqued, and conceptualized *a priori*. For instance, Epictetus proposes that we can pin down a particular role we assume and infer appropriate actions from such a role through logical reasoning (Johnson, 2014). In this sense, reason can also help us formulate reasonable defense for our role. In contrast, Confucianism takes a more empiricist stance on the essence of role. It argues that morality is not something that can be determined by preset standards of abstract roles but are determined "through particular interactions within the context of the roles that relevant persons live" (Ni, 2018, p. 194). In short, Western ethics treats role ethics as an ethical theory while Confucianism does not.

Western ethics and Confucianism adopt two different approaches to the relationship between non-institutional social roles (e.g., parent, sibling, friend) and institutional professional roles. For Western ethics, non-institutional roles and institutional roles need to be separated and they are different. According to Dare (2020), norms associated with institutional roles are determined by authoritative institutional design, whereas norms connected to non-institutional roles are generated by social practices around these roles in communities. Dare's famous "standard conception" of institutional roles argues that there is something distinctive about the operation of a professional institution that generates role differentiated obligations which may permit professionals to violate certain elements of ordinary morality (Dare, 2009). In contrast, Confucianism values the close connection between familial, non-institutional roles and public, institutional roles. More natural, familial, non-institutional role-based relationships (e.g., the parent-child relationship) can be extended to more public, institutional roles (e.g., the ruler and minister relationship). Non-institutional role-based relationships in the familial context such as the parent-child relationship often serves as a paradigmatic case for other relationships. For Confucianism, it is hard to believe that an unfilial child would be a loyal employee (or subordinate) (Chan, 2008).

Regarding the categories of roles, some early Western ethicists such as Epictetus do emphasize that there are some roles that are less social and are enabled by our role as rational beings. We automatically are given these roles as long as we are humans. Both Epictetus and Korsgaard would agree that there are certain fundamental roles that belong to us because of our rational identity (e.g., Epictetus's universal, cosmic role). For Confucianism, all roles start from the familial context. A somewhat extreme case for Confucianism is that responsibility would not exist if there were only one person in the world. Early Confucians such as Mencius would agree

that moral reasoning does play a critical role in moral decision-making. It is the capability to distinguish right from wrong that distinguishes humans from animals (*renqin zhibian*, 人禽之辨) (Wang, 2017). Nevertheless, in contrast to Western philosophers, Confucians also emphasize the critical value of emotional investment for demonstrating our commitment to the fulfillment of our communal roles and associated moral obligations (Chan, 2008).

With regard to role conflicts, Western scholars such as Korsgaard (2009) are heavily influenced by Kantian ethics and argues that our reason empowers us to freely renounce certain roles if these roles may cause a conflict among our duties. In comparison, from the Confucian perspective, Bell (2018) argues that we can appeal to moral standards external to our social roles and these external standards can help us better judge how to solve the issue of role conflicts.

6. Implications for the Design of Morally Capable Robots

In previous sections, we have discussed the synergies and differences between Western and Confucian perspectives to role ethics. As robot ethicists, our motivation for this analysis in part lies in our desire to develop new frameworks for analyzing the behavior of and informing the design of artificial moral agents, including morally capable robots. Accordingly, in this section we briefly discuss how the shared norms and irreducible differences between Western and Confucian approaches to role ethics can be helpful for reflecting on how we imagine and design morally capable robots.

We argue that role ethics provides a different approach to the anticipation about and design of future morally capable robots. Most previous work on artificial moral agents has been grounded in Western ethical theories such as deontology and consequentialism. Deontology in particular has attracted significant attention (Bringsjord, Arkoudas, & Bello, 2006; Scheutz, Malle, & Briggs, 2015) due to (1) moral psychological justification for grounding moral reasoning in application of bundles of deontic norms of prohibition, permission, and obligation (Malle 2017); and (2) the ready translation of logically encoded norms into natural language explanations (Kasenberg, Roque, Thielstrom, Chita-Tegmark, & Scheutz, 2019; Langley, 2019). However, due to the computational and philosophical limitations of purely normative and individualistic moral reasoning frameworks, researchers including us have recently begun exploring alternative possible theories that refocus on social-relational ontologies of social roles and relationships (Cappuccio, Sandoval, Mubin, Obaid, & Velonaki, 2021; Coeckelbergh, 2010; Kim, Wen, Zhu, Williams, & Phillips, 2021).

From the perspective of role ethics, a major task for the role ethics of technology is to investigate whether practices engendered by technology “are conducive or detrimental to our performance of the social roles” (Bell & Wang, 2020, p. 83). The development of robots can and should be encouraged by our political communities if robots help us realize our constitutive commitments or moral obligations prescribed by our social roles (e.g., child, parent, doctor). Similarly, robots that undermine the realization of our constitutive commitments should be restricted (Bell & Wang, 2020). However, to implement the role ethics approach in actual robot designs, there are some theoretical questions that need to be carefully considered.

The first question is what roles we can and should assign to robots. Both Western and Confucian approaches to role ethics acknowledge that the roles humans assume can be diverse and they include both more natural, personal social roles (e.g., friend, child, domestic partner) and more institutional and professional roles (e.g., doctor, engineer, nurse). For roboticists, can and should

we assign more natural, personal social roles, more institutional and professional roles, or both to robots? If we can assign both personal social roles and institutional roles to robots, are there any relationships between the two kinds of roles? In other words, can robots transfer their moral capabilities from their personal roles to institutional and professional roles or vice versa? In previous work (Williams, Zhu, Wen, & de Visser, 2020), we have primarily focused on the consideration of natural, personal social roles for robots. Because the traditional cardinal relationships such as those of Confucianism (father-son, husband-wife, older-younger, ruler-minister, and friend-friend) do not all apply clearly to robots (and in some cases impose sexist hierarchies we would not want to perpetuate in modern society), we have instead proposed new cardinal relationships for human-robot interaction, such as owner-ownee (an asymmetrical relationship that moreover only allows for robot embodiment of one involved role), adept-novice (an asymmetrical relationship that may change over time following changes in competence and experience), supervisor-subordinate (an asymmetrical relationship tightly related to organizational structure, which will likely tightly interact with institutional and professional roles), teammate-teammate (a symmetrical relationship that may similarly tightly interact with institutional and professional roles) and friend-friend (a symmetrical social relationship that may be of primary benefit and appropriateness in socially assistive robotics applications).

Second, as suggested by Epictetus and Korsgaard, humans can assume less social roles which are enabled by our identity as rational beings. Epictetus has further specified certain universal virtues that are related to our role as rational beings independent of any social context. Do we want to assign any role-virtues or role-norms to all robots regardless of their work context? Taking a pluralistic approach to role ethics might involve combination of computational systems for representing and reasoning with the relational roles described in the previous section with non-relational roles from the virtue ethics tradition, similar to the approaches recently proposed by Vallor (2016), by Kuipers (2018), and by Govindarajulu, Bringsjord, and Ghosh (2019).

Third, what normative power is attributed to robots based on the roles they assume, and how do humans infer and expect moral duties from the roles assigned to robots? Our comparative study of Western and Confucian role ethics has shown that both two traditions have acknowledged the descriptive aspect (e.g., what our roles are) and the normative aspect (e.g., what our roles morally require us to do) of roles. For future design of robotics, we argue that it is critical for all stakeholders including robots, human interactants, and designers to be aware of both descriptive and normative aspects of their own roles and those of roles assumed by others.

Robots need to be aware of their roles and associated moral obligations. Doing so allows robots to only perform tasks they are expected to do and generate explanations to justify their roles, obligations, and behaviors. Robots also need to be aware of the roles of human interactants. Doing so allows robots to have a contextual understanding of the real needs of humans and ensure that the decisions made by human interactants help fulfill their roles and moral obligations. Human interactants also need to understand robots' roles. Understanding robots' roles can be beneficial for more efficient and reliable human-robot interaction. It is helpful for humans to appropriately assess the moral consequences of robot decisions. Human interactants need to be aware of their own roles as well which will be helpful for them to critically examine to what extent their interactions with robots help fulfill their role-based moral obligations. Finally, robot designers can generate inspirations from their understanding of robots' roles. Their understanding of robots' roles can provide both descriptive and normative guidance on what functions and capacities should be integrated into robots.

Fourth, do we predetermine the moral obligations and standards for roles before we assign these roles to robots? Or do we let robots learn and determine moral obligations through their particular interactions within the context of the roles that they assume? Work from Kasenberg, Arnold, and Scheutz (2018) has highlighted significant challenges for approaches to automatically learning moral behavior from human data, such as through inverse reinforcement learning strategies. However, others, such as Sarathy, Scheutz, and Malle (2017), have provided principled human-guided approaches for learning deontic norms and weights thereon. Our analysis of Western and Confucian role ethics has shown that the two traditions have different takes on the essence of role. The Western tradition takes a more metaphysical stance on the essence of role. Roles and role-based obligations can be predetermined through our logical reasoning. In contrast, the Confucian tradition argues that roles are always socially constructed and are formulated in specific temporal and spatial contexts that constitute part of our everyday life. In designing morally competent robots, we suggest that it might be helpful to combine the two approaches. For instance, if Tina buys a robot who will become her companion at home, she expects that the robot has some preprogrammed moral obligations that are prescribed by the role of companion in a very general sense. However, once the robot joins the family, it will learn to become a good companion living with this *particular* person Tina who has unique life goals, needs, and habits. Tina's experience living with the robot will also be helpful for her to reflect on what her appropriate roles in relation to this *particular* robot. In such a process, we can imagine that some preprogrammed moral obligations of the robot will be further examined, refined, or even removed.

Finally, can roles themselves be sufficient for providing ethical and actionable guidance for human conduct? How do robots deal with role conflicts? Our comparative analysis of Western and Confucian approaches to role ethics has shown that scholars in the two traditions feel concerned about to what extent roles *per se* can provide clear, ethical, and realistic guidance for actions and whether there need to be more fundamental factors (e.g., standards, norms, principles) that guide our role fulfillment. For instance, to design a robot that serves the role of a caregiver, if the designer only relies on deriving design specifications from her reflections on the robot's role as a caregiver, it might be difficult to provide design solutions in situations that involve role conflicts. It is unclear what the robot caregiver is supposed to do when the patient refuses to take medication (e.g., whether the robot needs to prioritize its role in improving human wellbeing or its role in respecting human agency). Therefore, there needs to be some more fundamental moral principles that can help the robot make more realistic decisions in moral dilemmas. Even if we argue that robots should behave by considering what is the benevolent way to perform their social roles in relation to others, rather than considering whether actions are wrong due to violation of norms of prohibition, the judgment as to what actions are or are not considered benevolent with respect to certain relational roles will itself be grounded in norms encoding community consensus.

References

- Ames, R. (2021). *Human becomings: Theorizing persons for Confucian role ethics*. Albany, NY: SUNY Press.
- Ames, R., & Rosemont, H. (2011). Were the early Confucian virtuous? In C. Fraser, D. Robins & T. O'Leary (Eds.), *Ethics in early China: An anthology* (pp. 17-39). Hong Kong, China: Hong Kong University Press.
- Ames, R. T. (2011). *Confucian role ethics: A vocabulary*. Hong Kong, China: Chinese University Press.
- Angle, S. (2018). Building bridges to distinct shores: Pragmatic problems with Confucian role ethics. In J. Behuniak (Ed.), *Appreciating the Chinese difference: Engaging Roger T. Ames on methods, issues, and roles* (pp. 159-182). Albany, NY: SUNY Press.
- Bell, D. A. (2018). Roles, community, and morality: Comment on Confucian Role Ethics. In J. Behuniak (Ed.), *Appreciating the Chinese difference: Engaging Roger T. Ames on methods, issues, and roles* (pp. 203-212). Albany, NY: SUNY Press.
- Bringsjord, S., Arkoudas, K., & Bello, P. (2006). Toward a general logicist methodology for engineering ethically correct robots. *IEEE Intelligent Systems*, 21(4), 38-44.
- Cappuccio, M. L., Sandoval, E. B., Mubin, O., Obaid, M., & Velonaki, M. (2021). Can robots make us better humans?: Virtuous robotics and the good life with artificial agents. *International Journal of Social Robotics*, 7-22.
- Chan, J. (2008). Territorial boundaries and Confucianism. In D. A. Bell (Ed.), *Confucian political ethics* (pp. 61-84). Princeton, NJ: Princeton University Press.
- Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12, 209-221.
- Cottine, C. (2020). That's what friends are for: A Confucian perspective on the moral significance of friendship. In T. Dare, & C. Swanton (Eds.), *Perspectives in role ethics: Virtues, reasons, and obligation* (pp. 123-142). New York, NY: Routledge.
- Dare, T. (2009). *The counsel of rogues? A defence of the standard conception of the lawyer's role*. Farnham, UK: Ashgate.
- Dare, T. (2020). Roles all the way down. In T. Dare, & C. Swanton (Eds.), *Perspectives in role ethics: Virtues, reasons, and obligations* (pp. 31-44). New York: Routledge.
- Dare, T., & Swanton, C. (2020). Introduction. In T. Dare, & C. Swanton (Eds.), *Perspectives in role ethics: Virtues, reasons, and obligation* (pp. 1-7). New York, NY: Routledge.
- Ess, C. (2006). Ethical pluralism and global information ethics. *Ethics and Information Technology*, 8(4), 215-226.
- Govindarajulu, N. S., Bringsjord, S., Ghosh, R., & Sarathy, V. (2019). Toward the engineering of virtuous machines. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, Honolulu, HI. January 27-28.
- Hagen, K. (2010). The propriety of Confucius: A sense-of-ritual. *Asian Philosophy*, 20(1), 1-25.
- Higgins, K. M. (2018). Performance in Confucian role ethics. In J. Behuniak (Ed.), *Appreciating the Chinese difference: Engaging Roger T. Ames on methods, issues, and roles* (pp. 213-228). Albany, NY: SUNY Press.
- Johnson, B. E. (2014). *The role ethics of Epictetus: Stoicism in ordinary life*. Lanham, MD: Lexington Books.
- Kasenberg, D., Arnold, T., & Scheutz, M. (2018). Norms, rewards, and the intentional stance:

- Comparing machine learning approaches to ethical training. *Proceedings of 2018 AAAI/ACM Conference on AI, Ethics, and Society*, New Orleans, LA. February 2-3.
- Kasenberg, D., Roque, A., Thielstrom, R., Chita-Tegmark, M., & Scheutz, M. (2019). Generating justifications for norm-related agent decisions. *Proceedings of the 12th International Conference on Natural Language Generation*, Tokyo, Japan. October 29 – November 1.
- Kim, B., Wen, R., Zhu, Q., Williams, T., & Phillips, E. (2021). Robots as moral advisors: The effects of deontological, virtue, and Confucian ethics on encouraging honest behavior. *Proceedings of HRI '21: ACM/IEEE International Conference on Human-Robot Interaction (virtual)*. March 8-11.
- Korsgaard, C. M. (1993). The reasons we can share: An attack on the distinction between agent-relative and agent neutral values. *Social Philosophy and Policy*, 10(1), 24–51.
- Korsgaard, C. (2009). *Self-constitution: Agency, identity and integrity*. Oxford, UK: Oxford University Press.
- Kuipers, B. (2018). How can we trust a robot? *Communications of the ACM*, 61(3), 86-95.
- Langley, P. (2019). Explainable, normative, and justified agency. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, Honolulu, HI. January 27 – February 1.
- Luban, D. (1988). *Lawyers and justice: An ethical study*. Princeton, NJ: Princeton University Press.
- Malle, B. F., Scheutz, M., & Austerweil, J. L. (2017). Networks of social and moral norms in human and robot agents. In M. I. A. Ferreira, J. S. Sequeira, O. M. Tokhi, E. Kadar & G. S. Virk (Eds.), *A world with robots: International conference on robot ethics* (pp. 3-17). Cham, Switzerland: Springer.
- Ni, P. (2018). Does Confucianism need a metaphysical theory of human nature?: Reflections on Ames-Rosemont role ethics. In J. Behuniak (Ed.), *Appreciating the Chinese difference: Engaging Roger T. Ames on methods, issues, and roles* (pp. 183-202). Albany, NY: SUNY Press.
- Nuyen, A. T. (2007). Confucian ethics as role-based ethics. *International Philosophical Quarterly*, 47(3), 315-328.
- Pettigrove, G. (2020). Characters and roles. In T. Dare, & C. Swanton (Eds.), *Perspectives in role ethics: Virtues, reasons, and obligation* (pp. 11-30). New York, NY: Routledge.
- Puett, M., & Gross-Loh, C. (2016). *The path: What Chinese philosophers can teach us about the good life*. New York, NY: Simon & Schuster.
- Rosemont, H. (2015). *Against individualism: A Confucian rethinking of the foundations of morality, politics, family, and religion*. Lanham, MD: Lexington Books.
- Rosemont, H., & Ames, R. (2016). *Confucian role ethics: A moral vision for the 21st century?* Taipei, Taiwan: National Taiwan University Press.
- Sarathy, V., Scheutz, M., & Malle, B. (2017). Learning behavioral norms in uncertain and changing contexts. *Proceedings of the 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, Hungary. September 11-14.
- Scheutz, M., Malle, B., & Briggs, G. (2015). Towards morally sensitive action selection for autonomous social robots. *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Kobe, Japan. August 31 - September 4.
- Seddon, K. H. (n.d.). Epictetus. International encyclopedia of philosophy.

- <https://www.iep.utm.edu/epictetu/>. Accessed 12 April 2019.
- Swanton, C. (2020). Expertise and virtue in role ethics. In T. Dare & C. Swanton (Eds.), *Perspectives in role ethics: Virtues, reasons, and obligation* (pp. 45-71). New York, NY: Routledge.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. New York, NY: Oxford University Press.
- Wang, K. (2017). Xunzi's virtue ethics of rationality and the issue of emotions. In X. Yao (Ed.), *Reconceptualizing Confucian philosophy in the 21st century* (pp. 35-42). Singapore: Springer.
- Weinstock, D. M. (1998). Moral pluralism. In E. Craig (Ed.), *Routledge encyclopedia of philosophy*. Routledge. Retrieved January 5, 2021, from <https://www.rep.routledge.com/articles/thematic/moral-pluralism/v-1>
- Williams, T., Zhu, Q., Wen, R., & de Visser, E. (2020). The Confucian matador: Three defenses against the mechanical bull. *Proceedings of HRI '20: ACM/IEEE International Conference on Human-Robot Interaction*, Cambridge, UK. March 23-26.
- Zhu, Q., Williams, T., Jackson, B., & Wen, R. (2020). Blame-laden moral rebukes and the morally competent robot: A Confucian ethical perspective. *Science and Engineering Ethics*, 26(5), 2511-2526.
- Zhu, Q. (2018). Confucian ethics, ethical leadership, and engineering ethics education. *International Journal of Ethics Education*, 3(2), 169-179.