What Confucian Ethics Can Teach Us About Designing Caregiving Robots for Geriatric Patients

Alexis Elder

Short Abstract:

Caregiving robots are often lauded for their potential to assist with geriatric care. While seniors can be wise and mature, possessing valuable life experience, they can also present a variety of ethical challenges, from prevalence of racism and sexism, to troubled relationships, histories of abusive behavior, and aggression, mood swings and impulsive behavior associated with cognitive decline. I draw on Confucian ethics, especially the concept of filial piety, to address these issues. Confucian scholars have developed a rich set of theoretical resources for dealing with beloved but imperfect elders, and navigating the challenges of supporting seniors whose ethical commitments are unreliable. These resources provide a way to reconcile two important but conflicting desiderata: to value and care for seniors, but also to clear-mindedly deal with their moral shortcomings. In particular, they articulate a duty to *remonstrate* with our elders when they err. Confucian filial piety can helpfully inform robot design and use in geriatric care. They can be used to strengthen and protect emotional connections in important relationships, but should not be used to reinforce patient preferences when doing so damages relationships or their ability to act morally. Rather than conceive of patient wellbeing as in tension with moral behavior, and care as a burden for caregivers, not a source of value and meaning, Confucian accounts of filial piety help identify both new areas of concern and new potential in the development of caregiving technologies, ones which see these goods as complementary.

Introduction

Caregiving robots often assist with or offset human work in geriatric care, which means we ought to have a clear account of what such care consists in, and what geriatric patients are like. While members of older generations can be wise and moderate, possessing valuable life experience upon which to draw, or sometimes frail and vulnerable, in need of support in order to successfully exercise agency (Burema 2021), they also present a variety of ethical challenges, from prevalence of racism and sexism (Pew Research Center 2019) to histories of (and sometimes ongoing perpetuation of or complicity in) domestic violence (Gerino et al 2019), as

well as aggression, mood swings and impulsive behavior often associated with cognitive decline (Cipriani et al 2011). This dark side of aging and technology deserves more attention than it currently gets. How can or should these technologies account for elderly wrongdoing, and what special problems arise because of it?

In this project, I draw on Confucian ethics to develop a framework for addressing these issues, extending theoretical work in this tradition to serve concerns that are recognizable across many cultures and circumstances. Rather than restrict ethical frameworks to the cultural traditions in which they were developed, my aim here is to draw more widely applicable guidance, rooted in a philosophical tradition concerned with supporting human relationships and conditions for mutual flourishing. One aspect of Confucian ethics that I will argue is useful in geriatric care technologies is the idea of filial piety. In several of the classic Confucian texts, while filial piety is sometimes misconstrued as requiring unquestioning obedience of children to their parents, it is clear that Confucian thinkers consider this a misunderstanding, as we will see. Confucian scholars have developed a rich set of theoretical resources for dealing with beloved but imperfect elders, and navigating the challenges of caring for seniors whose ethical commitments are warped or simply nonexistent. Instead of taking patient wellbeing to be in tension with moral behavior, and instead of thinking of care as a burden for caregivers, Confucian accounts of filial piety help identify both new areas of concern and new potential in the development of caregiving technologies, ones which see patient wellbeing, value for caregivers in caregiving, and moral behavior of both seniors and their caregivers as complementary goals. These resources can organize and clarify a cluster of ethical issues around geriatric care as a distinctive locus of concern, moreso than in many ethical theories commonly deployed in Anglophone technology ethics. It is important to resist simplistic reductive accounts

of "East vs. West" philosophies in terms of "relational vs. individualist."[1] Nevertheless, one resource Confucian discussions of filial piety bring to bear is the importance of intergenerational relationships to our wellbeing, and the ways that our activities can nourish or injure them. While these relationships can be fraught, and members of both older and younger generations can err, they are worth attention and investment.[2]

These discussions provide a way to reconcile two important but conflicting desiderata: to value and care for seniors, but also to clear-mindedly deal with their moral shortcomings. In particular, they articulate a duty to *remonstrate* with our elders when they err, a duty that can helpfully inform technology design and use.

Robotic Design Challenges for Recalcitrant Seniors

In many parts of the world, increasingly "grey" populations require care, in societies in which younger generations face increased and competing workplace and social demands, leading many to invest hopes in social robotics to help, as the title of one paper puts it, "lift the burden" of geriatric care (Parks 2010). They may perform a variety of routine caregiving tasks, from medication management to companionship, assistance with household tasks, and activities of daily life like bathing, dressing, and fetching items around the home. This can complement or replace some efforts by both family members and professional caregiving staff. Such robots are likely to encounter elderly patients' behavioral issues, many of which seem to have a strong moral dimension. How should robots be designed to deal with these issues?

---

[1] See Olberding 2015 and Wong 2008 for some helpful discussions of how complex accounts of individuals, relationships, and social connections can be in Confucian ethics, and of course it would also be unfair to European or American philosophy to portray it as uniformly or unreflectively individualistic to the exclusion of social relations.

[2] While I focus here on interpersonal intergenerational relationships, scholarship such as Robert Neville's *The Good Is One, Its Manifestations Many* (2017) offers additional resources for thinking through the ways that Confucianism can accommodate larger-scale changes of institutions and social norms over time.

In order to better understand this, Draper and Sorell (2014) conducted a series of focus group discussions among elderly people, informal caregivers like spouses and family members, and formal carers from health and social services sectors. They discussed scenarios involving design of geriatric-care robots that were designed to highlight ethical issues. One involves a verbally abusive elderly woman whose behavior is alienating her daughter and interfering with her relationships with paid caregivers. It runs as follows:

> Nina, who is 70 years old, had a stroke two years ago but has now recovered the use of her arm, though one side of her face droops slightly. She is self-conscious about this, but it does not affect her physical functioning. She is supported at home by a Care-O-bot®. Since having the stroke she has become quite irritable and impatient. She often shouts at her daughter when she visits and complains angrily about her condition. Her daughter finds this very upsetting and has come to dread her visits. Nina has been so rude and demanding that two cleaners have already refused to work for her anymore. She is usually polite with her friends. Her Care-O-bot® has been programmed so that it will not do things for her if she asks sharply or in a demanding tone. It encourages her to say please and thank you and will withdraw help until she does so. Nina finds this infuriating and insists that the Care-O-bot® is reprogrammed to do what she asks no matter how she asks for help. (Draper and Sorell 2014, p. 126)

Note that in this description, the woman's irritability is connected to a recent stroke but does not impair her abilities in a uniform way: she is still capable of politeness with friends, but treats both robot and human assistants and junior family members badly. One focus group question was, should the robot be programmed to ignore her insults, comply with her, or require that she be polite to it?

Draper and Sorell note that responses to this scenario were mixed. Many, especially elderly people, felt it was unacceptable for a mere robot to try to control Nina's behavior. While sympathetic to the plight of her family members, they believed that the robot's role was to serve Nina regardless of her treatment of it. Focus group members were often drawn both to the idea that Nina couldn't help herself when being rude, and thus was not responsible for her actions, while also articulating the idea that the robot would be manipulative were it to try to shape her

behavior, thus violating her autonomy. Draper and Sorell point out that these ideas are in tension: "After all, if she could not control her rudeness, it is not an expression of her autonomy, and respect for her autonomy cannot therefore be used to justify toleration, especially given the apparently harmful effects of her rudeness on others," further noting that Nina's "daughter in particular was a captive of Nina's behavior, since, arguably, her filial obligations bound her more tightly than the obligations of Nina's cleaners bound them not to leave her employ" (Draper and Sorell 2014, p. 131), an important point that will be the centerpiece of a Confucian analysis in terms of filial piety. After criticizing arguments against programming the robot to interfere, they argue that the robot, while designed to serve elderly patients, should do so with the aim of facilitating her long-term rehabilitation, not gratification of passing whims. They draw on both consequentialist and deontological reasoning to defend their conclusion. In a consequentialist move, they point out that Nina is ultimately better served by a robot that discourages her antisocial behavior, which would pose a barrier to her re-entry to society and limit her long-term exercise of autonomy, as opposed to one that gives in to her preference for verbal abuse and reinforces bad habits that would interfere with her social functioning:

> Rehabilitation is meant in part to return a patient as far as possible to the health and independence they enjoyed prior to an adverse event. Against this background, it is reasonable for roboticists to design robots that can help patients like Nina to reconnect with social norms of cooperation by discouraging rudeness. (Draper and Sorell 2014, p. 131)

At the same time, given that patients are expected to have consented to therapeutic uses of robots in their rehabilitation and that their autonomy is enhanced by doing so, there seems to be deontological reason to favor their conclusion, as well, although noting limits to how much moral weight can be carried by consent:

The norms of rehabilitation rather than the older person's moment-to-moment wishes… govern robot-human interactions…. Its justification lies not just in the net benefits, but also in the minimization of harmful effects *and the agreement of the patient to both the ends and the means*. …But what if the older person lacked the capacity to agree to both the ends and the means? Would this make a moral difference? Not necessarily. The need for agreement to the ends and means is generated by respect for autonomy; and where capacity is lacking, so too is autonomy. (Draper and Sorell 2014, p. 132, emphasis added)

This scenario, although fictional, is (intentionally) useful for highlighting the complexity of ethical issues that arise in geriatric care. Patients' irritability and mood swings often target some more than others. Informal caregiving occurs after decades of established relationships, and statistics show that these include patterns that are not always harmonious. General population prevalence of domestic violence and child abuse (Huecker et al 2021), and less-legally-salient but ethically problematic interactions like habitual impoliteness and demandingness suggests that many seniors have imperfect histories with their caregiving networks, histories for which many bear at least some responsibility and which predate geriatric-onset cognitive impairments. Even without histories of explicit abuse, patterns of entitlement or expectations of deference may detrimentally impact interactions with caregivers, both paid/professional and unpaid family and friends. And people may often find themselves caught between excusing current misbehavior as involuntary, and the temptation to let it go out of respect for autonomy, despite the fact that one reason undermines the other.

One point I wish to draw our attention to is that geriatric caregiving poses special concerns because caring for one's spouse or parent can include relational frictions arising from historical patterns and concerns. For instance, although some participants argued that Nina should be free to abuse the robot because it feels nothing, some users of household virtual assistants are noting that a family member's (mis)treatment of a robot can be disturbing to others in the household, because they are already sensitized, through family history, to raised voices,

verbal abuse, and so on (Dreyfuss 2018). These issues should be expected to be exacerbated in geriatric care robotics, because they are positioned to assist family members in caregiving, sometimes allowing seniors to delay entry into professional assisted-care settings, and can in principle harmlessly absorb ill-treatment patients more typically direct toward caregivers.

These concerns can be exacerbated when robots are designed to adapt to individual patients' preferences. This worry has been dramatized in the film *Robot & Frank*, a near-future story about a geriatric-care robot assigned to look after an old man with a history as a jewel thief. Adapting its activities to suit the interests of its patient, the robot is eventually drafted to participate in a jewel heist, enabling the man to relive his glory days but further estranging him from his family, who suffered consequences of his earlier criminality and resent its resurrection. More generally, designers may need to ask when and whether one ought to adapt to seniors' current patterns of interaction, or whether instead one ought to engineer these robots to resist contributing to, or even actively working to change, problematic behavior patterns, even against the wishes of their charges. Given recent trends toward personalization of technologies that are facilitated by developments in machine learning, if such programs are designed to adapt to patients' patterns of interaction and language use, they may learn to accommodate or even reproduce abusive, racist, sexist, homophobic or otherwise problematic behavior of the seniors to which they are assigned.

Draper and Sorell's response to this sort of concern endorses a kind of therapeutic paternalism. Either patients are not acting voluntarily when they behave abusively, and therefore paternalistic intervention in the interest of therapeutic goals is justified, or else their behavior is voluntary but we reject the idea that this justifies protecting it when it interferes with the patient's interest in social integration. But such a paternalist response might not be justified, or at least

might need more argument. For example, intervention might be ineffective, if it involves swimming upstream against deep-seated patterns of behavior. (If Nina has an entrenched habit of treating her daughter and subordinates badly, or suffers short-term memory impairment, it may not be reasonable to expect much therapeutic benefit from correcting her.) It might also be inapt for patients whose long-term prospects preclude much engagement with the broader social world, as during hospice care. And furthermore, one might question whether an autonomy-centered framing is the best way to address seniors' misbehavior. There seems to be something to the idea that their social interactions and relationships are part of seniors' big-picture wellbeing, but I will explore a different way to proceed, one that does not hinge on long-term prospects for exercising agency in broader society, but instead focuses on supporting established and valuable relationships with children and caregivers, ones which are plausibly partly constitutive of their ability to live well, even in situations where they may not be able to govern their own behavior without external support.

Confucian Ethics and Filial Piety

In this section, I connect the concerns we have just seen with Confucian accounts of how filial piety directs us to respond to wrongdoing by our elderly relatives. Taking filial piety as a foundation for addressing such concerns can be helpfully contrasted with Draper and Sorell's approach in that, unlike many European-tradition conceptions of paternalism, which take parent-child relationships to paradigmatically involve parents knowing what is best, making paternalistic geriatric care a kind of inversion that casts the parent in the role of the child, sophisticated accounts of filial piety are directly intended to provide guidance for elderly parent-adult child relationships. This will help better explicate the features of Confucian filial

piety that can be useful in designing geriatric care technologies with an eye to addressing geriatric wrongdoing.

To recap, the picture that we have seen is one in which seniors err morally, and where their position *as* seniors contributes to concerns about how to intervene: both in expectations of deference and resistance to correction by juniors. It would be problematic to leave them to take care of themselves, because they are in many ways ill-equipped to do so, but at the same time their behavior can harm others and thereby (given the importance of interpersonal relationships to their support networks) harm themselves and cannot merely be bemusedly tolerated. Treating them either straightforwardly paternalistically, or as autonomous adults capable of and responsible for managing their own affairs, seem unsatisfactory. Where, then, can we turn for guidance?

Pretheoretically, it would seem that the outcome at which we aim with good geriatric caregiving should include maintaining valuable intergenerational bonds and avoiding the harms that are associated with geriatric wrongdoing, although as will be discussed, understanding what these harms consist in will require some work. Attention to the structures and practices in which these technologies are used will be important to address the wrongdoings relevant to them in particular. For these reasons, Confucian ethics seems to offer great promise.

Background

First, some background may be helpful. Commonly understood as a variety of virtue theory but also a kind of role ethics (Ames 2011), it emphasizes interpersonal relationships, argues that identities and moral principles are socially embedded and reinforced, and includes a detailed account of moral psychology and practices that facilitate growth and sensitivity to social

placement[3]. While criticized by contemporary philosophers of the Mohist school for prioritizing care for family over impartial or universal concern for all, scholars such as Mengzi defended the practice on the grounds of its psychological impact in developing our capacity to care, resulting in an exceptionally detailed account of the affective components of caring as developed in interpersonal relationships. Mengzi argued that the Mohist challenge can be met, by focusing on how to 'extend' these emotional elements of care to promote political institutions that support the wellbeing of the elderly and vulnerable in general (Pang-White 2011). Furthermore, its emphasis on *li*, translated as both "etiquette" and "ritual", provides rich resources for understanding patterned interpersonal interactions within broader implicit social structures, as in Amy Olberding's discussions of the wrong of rudeness (Olberding 2019). It thus seems well-suited to the problems identified above.

Confucian ethical theory assigns a central role to filial piety, the duty to care for one's parents. While different philosophers within the tradition interpret this differently (as one might expect), I am interested guidance about how caregiver should respond to geriatric wrongdoing. Thus, I review several arguments that this concept, despite sometimes being misunderstood as a duty to do what one's parents say uncomplainingly, or to accept mistreatment, actually contains rich theoretical resources for resisting seniors' misbehavior without abandoning them. Because adult children often end up playing a role in caregiving for geriatric patients, in some cases this framework can be directly applied. However, I also think that it can yield insights into the particular challenges of caregiving for geriatric patients, and relationships between caregivers and their elderly patients more generally, and in what follows will extend filial concerns to intergenerational caregiving more broadly.

---

[3] The extent to which this is compatible with Western conceptions of care ethics is the subject of robust debate by Confucian scholars (Li 1994, Rosemont 1997, Herr 2003, Lambert 2016) and I do not intend to take sides here.

Protecting Affective Ties Necessary for Relationship

One classic example of how filial piety can involve a duty to care for one's parents without submitting to them is found in the story of the legendary sage-king Shun, upheld as an exemplar of virtue in core Confucian texts. Shun's parents were notoriously abusive and bent on destroying Shun's happiness at every opportunity, making his task of caring for them quite difficult. But his ingenious solutions were discussed in the *Mencius* as illustrating an ideal of how to exercise filial piety in difficult circumstances. David Wong's summary of this dynamic is instructive.

> When the time came for Shun to marry, he knew that his father would refuse permission if asked. So Shun did not ask [but married without their permission]. One reason given for this surprising decision, coming from the ultimate filial son, is that the worst way of being a bad son is to provide no heir (4A26). It is the other reason given, however, that I want to highlight. A man and woman living together, it is said in the *Mencius*, is the most important of human relationships, and if Shun had set aside that most important of relationships, it would have caused bitterness toward his parents (5A2). The reason it would have caused bitterness, of course, is that one of Shun's most urgent of interests would have been frustrated. Its satisfaction is crucial for the viability of his relationship to his parents. It would have been foolish for Shun simply to have swallowed his bitterness and submitted to what he knew his parents' wishes to be, foolish in terms of his own interests in marriage, and foolish for his relationship to his parents. (Wong, 2004, pp. 427–428)

That is, caring for parents need not be at odds with caring for oneself or mitigating the harms they are capable of doing to others. Instead, part of caring for them can include preserving the conditions necessary for them to enjoy good relationships with their caregivers, to the point of acting against their expressed or anticipated preferences and in favor of their real interests. Shun needed to protect his capacity to care for them without bitterness, which meant not putting himself in a position where their treatment of him would produce such an effect.

While this might look quite a bit like Sorell and Draper's defense of the therapeutic correction of rudeness by caregiving robots, it is subtler in focusing not just on enhancing a patient's long-term autonomy, or ability to function in "society" writ large, but in its focus on the particular harms done to interpersonal relationships with caregivers, and the importance of preserving, insofar as it is possible, the affective ties between particular people that are so important to the function of families and the possibility of a human life lived well. If this is consequentialist, it is a very different form than that of European ethical traditions, one in which individual psychological dispositions, interpersonal relationships, and occupations of social roles are both conceptually and instrumentally intertwined.[4] To the extent that even smaller cases of seniors' misbehavior, like the ones in Sorell and Draper's scenario, harms their relationships with their younger kin, it would seem that Confucian accounts of filial piety have something to say about it.

Protecting Moral Uprightness (One's Own and One's Parent's)

Further exploration of this complex notion of filial piety is helpful. After all, one might think that family relationships can also be preserved by accepting one's parents despite misbehavior, and helping to shield them from the consequences of their actions, and that where possible this would be on a par or even preferable. This is in fact suggested by some passages in the *Analects*, such as a story in which someone brags about their province being so committed to morality that a son would turn his father in to the authorities for stealing a sheep, only to be criticized by Confucius for holding this as the relevant standard for moral conduct. Confucius says, "In my village, an upright person is different; father does not disclose son's wrongdoing,

---

[4] For an extensive discussion of the extent to which Confucian ethics does and does not resemble other consequentialist frameworks, see Ivanhoe 1991.

and son does not disclose father's wrongdoing, and the [moral] uprightness is in it" (*Analects* 13.18)

However, scholars such as Huang (2013) have argued that even in this case, we ought not read Confucius as advocating for shielding parents' misbehavior, but rather for creating optimal conditions for correcting parents' misbehavior, carefully considering relational and psychological aspects of remonstrating with one's parents effectively. His argument runs as follows.

Confucius clearly and repeatedly emphasizes that filial piety does not consist merely in physical caretaking for parents' external needs. We have to be careful not to confuse (mere) physical care with emotionally engaged and respectful care. Genuine fulfillment of filial piety requires moral maturity on the part of the adult child caregiver. This point is articulated in a classic treatise on filial piety:

> In the *Book of Filial Piety*, it is … claimed that a filial person "does not do anything against moral principles… Thus, even if one's words spread through the whole world, there is never any fault; and even if one's action affects the whole world, there is never any complaint" (*Xiaojing* 4).... (Huang 2013, 130)

It is not merely the case that exercising filial piety is part of being moral. Being moral is integral to exercising filial piety. For people trying to deal with problematic parents, catering to their desires when those desires are immoral is not part of genuine care, in part because it implicates one in their wrongdoing, making their children less of a source of merited pride, and introducing a potential (warranted) cause for shame. This, then, can be a distinctive source of moral authority for children in relationships with their parents, authorizing them to reject their parents' subjective values in favor of doing what is genuinely moral. This point is made explicit by Confucian scholar Xunzi:

Xunzi summarizes that "there are three scenarios in which filial children ought not to obey their parents: (1) if their obedience will endanger their parents, while their disobedience will make their parents safe, then it is truly loyal for filial children to not obey their parents; (2) If obedience will bring disgrace to their parents, while disobedience will bring honor to their parents, then it is moral for filial sons to disobey their parents; (3) if obedience will lead to a life of beast, while disobedience will lead to a civilized life, then it is reverent for filial children to disobey their parents. Therefore, it is not proper for a son to not obey what should be obeyed, and it is not loyal for a child to obey what cannot be obeyed. It is great filial piety to understand when to obey and when not to obey in order to be reverent and respectful, loyal and trustworthy, and act with sincerity and carefulness" (*Xunzi* 29.2) (Qtd in Huang 2013, 132)

In fact, there is robust textual support for the idea that for Confucius, one's filial duty is supposed to include not just being moral oneself (and not letting one's parents corrupt one) but also an obligation to remonstrate with parents when parents go astray. This is well-illustrated in the following passage from the *Konzi Jiayu*, a collection of stories about and sayings of Confucius himself:

"Zigong, one of Confucius's students, wants to confirm with Confucius that one's obedience to parents is filial piety, just like a minister's obedience to the king is loyalty. However, Confucius replies, "How shallow you are! You don't understand. In the ancient [sic], when a good king of a big state has seven ministers who dare to remonstrate, the king will not make mistakes; if a middle sized state has five remonstrating ministers, the state will have no danger; if a small state has three remonstrating ministers, the official salaries and positions can last. If a father has a remonstrating child, he will not fall into doing things without propriety; and if a scholar has a remonstrating friend, he will not do immoral things. So how can a son who merely obeys parents be regarded as being filial, and a minister who merely obeys the ruler be regarded as being loyal? To be filial and loyal is to examine what is to follow (*Kongzi Jiayu* 9; 57; the same passage with slight variance also appears in *Xunzi* 29.3)" (Qtd in Huang 2013, 131-2)

Effective, Relationship-Preserving Remonstration

It would be too quick, however, to move immediately from this point to the conclusion that children as opposed to caregiving robots are the best ones to do the remonstrating. Huang argues that the determining factor guiding when – and from whom – remonstration is delivered, is what intervention will actually lead to change, or have the highest likelihood of doing so,

given the parent/child relationship and the importance there of trust, respect, and mutual affection. At the *Analects* 4.18 it is said that "when serving your parents, [if they are wrong] you ought to **gently** remonstrate with them" and… "as is stated in the *Book of Rites*, "one ought to remonstrate with low tone, nice facial expression, and soft voice" (*Liji* 12.15) (Qtd in Huang 2013, 133) There is some dispute over how to translate the second half of Analect 4.18, with some arguing that it says that, roughly, that if parents are unresponsive to remonstration, children ought not go against them. But Huang argues (along with others) that the thing children ought not go against is their own will; while they should "remain reverent toward" their parents, they should not "go against" their own will (that is, they should persist in their remonstrations). (133-135) Note that even if Huang's interpretation is rejected, on the alternative understanding it is still worthwhile to try to remonstrate, and desirable to do so in a way that is both gentle, respectful, and effective; the interpretive dispute is over how long one should persist. If robots prove to be more effective, or better at preserving affective ties between parents and children, in at least some circumstances, one would still have reason to use robots in these contexts.

Huang notes that the literature among Confucian scholars includes extended discussion of how best one ought to remonstrate, a conversation that clearly involves both moral psychology and context-sensitivity. What is clear, however, is that one's commitment to moral principles does not diminish in the face of parental wrongdoing, and in fact that filial piety requires that one extend one's concern to moral principles to include their misbehavior as well as your own.

Huang points out that one passage from the *Book of Rites* seems to suggest that you even ought to put yourself in harm's way to remonstrate with immoral parents: "if your remonstration is not taken by your parents, you ought to remain reverant and filial. If they are happy, you ought to resume gentle remonstration; if they are not happy, however, instead of letting parents cause

harm to neighbors, you ought to use extreme form of remonstration. If at this extreme form of

remonstration your parents get angry and unhappy, hitting you hard with whips, you still ought

not to complain against them; instead you ought to remain reverent and filial to them" (*Liji*

12.15) (Qtd in Huang 2013, 135) But he points out that this is inconsistent with Confucius'

arguments that, as we saw above in the discussion of King Shun, one ought not let one's parents

abuse oneself in a way that would damage the affective ties between parent and child.

"Confucius disapproves of Zengzi's blind obedience [which leads him to let his parents strike

him] and asks him to follow the example of the Sage King Shun, who does not let his parents

commit wrongdoing toward him." (Huang 2013, p 138)

Huang concludes that the most plausible way to read the sheep-stealing discussion noted

above, which seems to support protecting parents from warranted punishment, is not to conceal

one's parents' wrongdoing, but to play a role in its correction that leaves open the trust and

emotional relationship that allows one to effectively remonstrate. Later discussion of King Shun

makes clear that when his parents err, he ought not turn them to the judicial authories himself but

also ought not use his position as ruler to prevent them from being convicted. Instead, while the

law ought to prosecute law-breakers, the role of the child in such situations is to use the unique

family ties to correct the erring family member as gently but effectively as possible, without

endangering themselves. Huang concludes, "While a filial person, by definition, ought to take

care of his or her parents, for Confucius, this not only includes parents' external well-being but

more importantly their internal well-being. Thus it is extremely important for filial children to do

all that they can to make sure that their parents do not commit wrongdoings that will cause

damage not only to the external well-being of the victims of their wrongdoings but also to their

own internal well-being." (Huang 2013, p. 149)

Facilitating Moral Agency

This enriches our understanding of what is at stake when geriatric patients err. It is not merely that when they are rude or abusive to caregivers, or that they practice habits that will make it harder for them to get along with others in society. They may behave in ways that make them worse people, even by their own lights, and damage their loved ones and their valuable personal relationships. Part of protecting them and caring for them can include caring for their own integrity and the wellbeing of their children, as well as their relationships with their loved ones. This is not an individualistic conception of moral agency, but rather one where a person's capacity to do things can require contextual and social supports: my agency as a driver is shaped by, among other things, the car, the roads, and the behavior of other drivers. Facilitating seniors' moral agency can be valuable even in situations where "reintegrating with society" is not the main goal, and shows where therapeutic goals can overlap with palliative ones. Being a good person, having children and relationships that are thriving, matter intrinsically and do not depend on achievement of further goals for their significance, as in Sorell and Draper's instrumentalist reading of the harms of rudeness and abusiveness.

What A Duty to Remonstrate Means for Technology Design and Use

When it comes to rudeness and verbal abuse by geriatric patients, as we have seen, the situation is complicated by the fact that caregiving relationships are often already fraught with personal history. Robots may thus be useful in a Confucian framework to help caregivers care for themselves and simultaneously protect their affective ties to the patient, sometimes by getting some much-needed distance and respite, in order to preserve rather than escape their caregiving relationships, given some seniors' tendencies to lash out at caregivers and those close to them. But robots can also be useful in shaping what Zhu et al call a "moral ecology", supporting

seniors' moral agency and promoting moral uprightness both by modeling moral and polite behavior themselves and by issuing rebukes to human beings engaging in rude or abusive behavior (Zhu et al 2020). And this is a framework that can be useful both within and outside of cultural contexts traditionally associated with Confucianism. The arguments I present do not require any pre-existing commitments to a particular cultural paradigm, but follow from much more widespread concerns about geriatric wrongdoing and the importance to children and family members of preserving affection and trust without overlooking moral concerns. This is not, of course, to deny the importance of culture to how ethical concerns are understood or made manifest. Rather, it is to show how a philosophical framework developed in one cultural context can shed light on cross-cultural concerns about elder care.

As we have seen, the issue of distinguishing parental care from obedience to parents was important enough that Xunzi explicitly offers three main reasons to disobey a parent's orders within the framework of filial piety: safety, honor, and civility. More generally, we might identify both procedural concerns and specific points of focus which might inform technology design for geriatric care.

Protecting Moral Uprightness

When it comes to procedural concerns, the following advice emerges: First, it is not just possible but sometimes necessary for care to be administered without submitting to instructions. As was noted earlier, "To be filial and loyal is to examine what is to follow" (*Kongzi Jiayu*), checking to see whether someone's immediate activities actually promote their values. Specific points to examine in the interest of promoting values involve the structures and practices in which caregiving technologies are found, from ways that people's identities emerge from the roles they occupy for each other (parent, child, neighbor, citizen) and that specific obligations

attached to these roles can be of major significance to people's wellbeing as occupiers of these roles (so being a good parent is part of being a good person for those who occupy the role of parent). For example, a caregiving robot might insist upon civility in interactions for which it is present, responding only to instructions phrased politely, as in Sorell and Draper's example, or even issue verbal rebukes during rude interactions, whether directed at the robot or at those around it. The robot might, for example, respond "that's not nice!" to insults and invectives, providing reinforcement to human caregivers should the patient target them as well. Rather than aim at frictionless fulfillment of the patient's desires in an attempt to maximize pleasure and minimize discomfort, as a utilitarian framework would suggest, or issue rebukes only when and because it facilitates a patient's therapeutic goals contingent upon reentry to society, the robot might be designed to disrupt harmful patterns of engagement and help protect (within reason) human caregivers. This may also benefit children and other caregivers by creating or facilitating contexts in which it is more feasible for them to respond to seniors' needs with emotional sensitivity and compassion. By helping them to behave virtuously, this contributes to the wellbeing of the parent/child relationship by helping them to be virtuous children (thus fulfilling an important part of their roles *qua* child) and to nurture the relationship.[5] This leads us naturally to the next concern.

Promoting Harmonious Interactions

Second, the way morals are socially embedded and not fully up to individuals: protecting people from opportunities to misbehave and providing them with opportunities to act well can be part of caring for each other and, in turn, supporting relationships. If presenting a caregiving robot as feminine increases the likelihood of the person's directing gendered abusive language

---

[5] I thank a referee at this journal for suggesting this point.

toward it, for instance, as has been found in interactions with feminized virtual assistants (West et al 2019), this may be a reason to avoid designing assistive robots to activate gender schemas, or at least feminine ones, even if there are minor utility gains to be realized by tapping into established expectations around feminized caregiving work that make patients somewhat more comfortable with receiving care from robots perceived as feminine. This, again, differs from frameworks that focus on maximizing subjective user comfort, at least in the absence of clear causal contributions to perceptible harms that outweigh the benefits, as well as autonomy-based framings that focus on promoting the patient's capacity to exercise choice rather than protect them from ill-considered use of this capacity to undermine their own values.

And finally, that these two issues of prioritizing care over obedience and structuring interactions to facilitate better behavior can overlap. When robots are designed to help patients to act well in established caregiving relationships, they help the patient, because protecting a relationship can protect identities constructed through that relationship *and* the trust and affective dispositions that support care. That is, a rich conception for care of patients should include care for their moral agency and behavior. At the same time, finding ways for robots to help human caregivers provide care with appropriately caring and respectful attitudes can be an important goal to keep in mind as these technologies are developed.

Effective Remonstration

In order to design technologies to support elder care consistent with filial piety, designers will need to attend to moral psychology: what presents temptation to act badly? What facilitates moral growth, or helps sustain moral activity, as well as dispositions to care, on the part of both patients and their loved ones? And how does social placement matter, in a given situation? Is some remonstration appropriate to a relationship (for example, is the issue something that should

be addressed by a child, or a spouse, or a friend) or is addressing that issue potentially disruptive to the particular relationship and better outsourced to a third party, like a medical provider, administrator, or even legal representative? These factors direct our attention to both features of the design process, and particular goals in implementation. A full accounting will need to include the ways that these issues implicate cultural practices and thus can be expected to vary by cultural context, itself a complex issue (Dennis and Clancy 2022). But there are some points we can identify as worthy of attention, while bearing in mind that their implementation may look different in different contexts. I begin by surveying process issues, before articulating specific goals in ethical use of geriatric care technologies given the preceding discussion of filial piety.

Context-Sensitivity

In terms of process, first and foremost is the point that ethical remonstration in cases of geriatric wrongdoing will require situational and psychological sensitivity - design goals should include modularity rather than assuming one system will fit all users. In some cases, for example, remonstration by family and friends may be more effective than depersonalized robotic intervention. (This may be one reason conflicting intuitions were so prominent in the case described by Sorell and Draper - participants might be imagining that remonstration by Nina's daughter would be more effective than a patronizing robot.) But as Confucius argued, making oneself a target of one's parents or elders' wrongdoing is not what filial piety calls for, and in fact enables the parent's wrongdoing. And some people may not be able to remonstrate, at any rate not on certain issues, and/or may benefit from reinforcements, including by artifacts. Robotic reinforcement of social norms may be appropriate, to the extent feasible within the limits of particular interpersonal relationships, and may be appropriately partnered with direct personal interventions by caregivers, with the ultimate aim of preserving and supporting

interpersonal relationships, and the details of the particular relationship and its history can change what kind of intervention is both effective and appropriate. In what follows, I will give examples of what might be considered appropriate interventions by technologies, ones that draw on, reinforce, and protect patients' relationships with their children and caretakers.

Preventing Harm to Relationships

In design and use, one goal that emerges is to prevent harms to both self, others, and existing relationships. While this might sound obvious, our understanding of harm is enriched by the discussion of filial piety found in the literature, as well as how it can be caused by geriatric wrongdoing. For example, it turns out to be important in caregiving to avoid letting parents cause bitterness in their children and caregivers, in order to preserve childrens' psychological dispositions necessary to have appropriate affection and concern for their parents. This means children and caregivers should not be put in a position where parents' (or other geriatric patients') mistreatment would produce relationship-impairing psychological damage to them, and steps should be taken to preserve affective ties to the patient. This is a complex interaction of affective and emotional labor, one which, with thoughtful planning, may include social robotics, albeit with attention to how they fit into the broader emotional dynamics of the situation. (Dobrosovestnova et al 2021) One might object that this is already a feature of any mature moral theory; one ought not help the patient hurt others. But this account differs from a general injunction against harming others, in that it directs attention to particular harms done to specific interpersonal relationships, which may be vulnerable in different ways depending on both interpersonal history and individual psychological dispositions. While a less nuanced account of harm-avoidance invites one to think of efforts to prevent harms in terms of tradeoffs

(is some instance of harm "worth it" in terms of net harm vs. wellbeing-promotion?) this approach reframes the issue such that a child or caregiver has more resources to resist feeling "caught" between caring for their parent, and protecting their own integrity or wellbeing: it damages the parent-child relationship for the child to debase themselves, neglect themselves or betray their own integrity, by making the parent complicit in being worse *as* a parent (even if the parent is not in a position to recognize this). The goal of caregiving shifts to protecting the geriatric patient in a way that protects and promotes their established relationships as well as possible, given the constraints.

In one relatively small-scale (insofar as night-time wandering is a relatively innocuous example of 'wrongdoing') but instructive example of re-imagining care for the elderly in a way that protects relationships, Sharkey and Sharkey (2012) describe a system to care for an elderly man with dementia, constructed in partnership with the man's daughter. "A system was developed that helped with his disorientation and confusion at night—when he left the bed at night, the light in the toilet was turned on, and a voice prompted him to visit the toilet. Similarly, if he went near the door of the apartment at night, the recorded voice of his daughter was played, encouraging him to go back to bed. If he ignored this, and left the apartment, then care staff in the building were alerted." (Sharkey and Sharkey 2012, p. 33) In this way, the man's existing trust and responsiveness to his daughter were leveraged to his benefit, as was her concern for him, by helping her to reassure him and redirect him when he wandered, without the kinds of sleep disruptions that can prove very taxing to caregivers for such patients. (Leggett et al 2017)

Protecting Patients' Moral Agency

This injunction against letting patients harm caregivers involves another goal, as well, not just in terms of damage to caregivers' relationships (or physical concern for caregivers, although

23

this is of course important as well) but from a concern about the harm to patients of putting them in situations in which they "commit wrongdoing". As we saw in the discussion in the previous section, filial piety involves a responsibility to prevent parents from making mistakes, keep them from "doing things without propriety", and more generally keep them from doing immoral things.

One ought to, on this account, protect parents' internal wellbeing, understood as their moral character, as well as that of potential victims. While an atomic individualist conception of agency would focus on separating out and respecting parameters for autonomous action, the relational approach found here emphasizes providing people with frameworks in which it is easier to act well, and preventing them from ending up in situations in which they are overcome by the temptation to act wrongly. This can have a variety of implications for technological design and use. For example, the developers of the therapeutic robot seal Paro found that it could be beneficial in reducing aggressiveness and increasing prosocial behavior among the elderly, illustrating one way in which technologies can help create conditions in which it is easier for patients to act well. But even above and beyond patient discomfort associated with anger and agitation, an interest in promoting their internal wellbeing can include not setting them up to fail in other ways less immediately linked to patient discomfort.

One might be tempted to argue, at this point, that patient comfort ought to take priority over enforcing social norms, etiquette, and policing of other 'small' offenses. But this overlooks the importance of framing situations so as to facilitate patients' abilities to act well, and to avoid presenting them with temptations to act badly when they are likely to do so. As noted earlier, gender can play a role in how patients respond to technologies, a phenomenon that is not limited to geriatric care technologies. Many current personal digital assistants like Siri and Alexa have

been default 'gendered' feminine, a design decision defended in terms of user acceptability studies by developers (Hempel 2015). But over and above concerns about whether this *reinforces* or *amplifies* sexism, it can offer opportunities to engage in sexist verbal abuse without pushback when it fails to operate as expected (Dreyfuss 2018) and this may itself seem like a kind of harm to geriatric patients *even when their gender stereotypes are already set*. The goal of gender-neutral or masculine caregiving robots is not, on this account, to undo these embedded expectations about gender, but given these limitations, to avoid activating them in ways that invite the patient to act worse than they would otherwise. That is, by providing them with situations where they find it tempting to engage in sexist verbal abuse, a Confucian account of filial piety would find this to be a situation where caregivers go wrong in putting the elderly in positions where they are likely to "do things without propriety", setting them up to behave badly. Caregivers and designers ought not to make it actively harder for geriatric patients to act well, and this has implications for how gender, race, and ethnic schemas may be activated by robot design features.

Protecting Relational Capacity for Remonstration

One other feature that emerges from this account of filial piety is the importance of protecting the patient's trust in and affection for family members and caregivers, something that can affect their long-term openness to remonstration in the context of interpersonal relationships. This ties into the discussion in the last section about the possibility of programming robots to intercede in defense of caregivers when patients are rude or verbally abusive. It is consistent with filial piety to hold that in some cases, a perceived neutral third party (equivalent to "the law" in Confucius' discussion of sheep-stealing) may be better suited to monitoring and/or reprimanding wrongful activities. Zhu et al's discussion of a Confucian role for robotic moral rebukes offers a

detailed survey of psychological and empirical research that can be useful in developing

appropriate technologies. (Zhu et al 2021) In particular, it can be important that some behaviors

be addressed without creating friction in families and caregiving relationships by creating

mechanisms for enforcement that do not depend on the patient being "turned in" by their

children or caregivers. This partnership can be an important aspect of creating optimal conditions

for correcting parents' misbehavior. For example, if the patient is recalcitrant about taking

medication (or engaging in physical therapy, or other perceived tedious or unpleasant but

medically important activity), it may be better to have a robot administer medication reminders

and keep a log to be shared with medical providers, rather than turning medication monitoring

into an interpersonal struggle with family or primary caregivers, one in which loved ones are

perceived to have 'turned against' the patient. More generally, when using ideals of filial piety as

guidance in developing and using geriatric care technologies, it is important to consider both

relational features (who is doing the correcting, and how will this affect the relationship?) and

psychological ones (what situations/features of presentations make the patient receptive to

remonstration?).

Lessons Learned

Therapeutic robotic interventions offer promise both in repairing the foundations of

interpersonal relationships that can be used to engage in more mutually respectful and

affectionate remonstration, although they should not be used as a substitute for interpersonal care

unless doing so is necessary to protect family members and caregivers from harm. Concerns

about protecting elderly "autonomy" by providing abusive behavior with targets designed to be

subservient fail to promote the interests of either the patients or their families and social circles.

By helping to offload difficult initial exchanges about wrongdoing, they may create new

opportunities for interpersonal intervention that promote wellbeing of both seniors and their caregivers. And, importantly, "user acceptability" initiatives should not reinforce patients' tendencies toward wrongdoing.

As we have seen, older people can engage in moral wrongdoing that involves a variety of relational features, from presumptions of wisdom or fear of conflict by younger family members, to presumptions of entitlement of deference that can amplify these problems. Confucian ethics, with its focus on cooperative flourishing and the psychological processes necessary for moral development, provides resources for thinking about these issues in context.

There may be room for a divided approach: while it might not be beneficial for junior social connections to "call down the law" on misbehaving seniors, for fear of destroying what special capacity for remonstration they have in virtue of interpersonal trust, this does not rule out impersonal quasi-legal intervention as an additional tool to manage geriatric wrongdoing. A blended approach that includes policies that effectively curtail misbehavior, as well as concerted courageous intervention by younger family and friends to remonstrate with the elderly, not as part of public humiliation (which may destroy the trust and affection that makes them effective remonstrators) but as an expression of concern, may be an important strategy to pursue.

Remonstration by family and friends may be more effective than depersonalized robotic intervention, but more likely to be called for when family and friends are the targets of the wrongdoing, and as Confucius argued, making oneself a target of one's parents or elders is not what moral principles call for, and in fact enables the parent's wrongdoing by providing a handy outlet. To this end, robotic reinforcement of social norms may be appropriate, but to the extent feasible within the limits of particular interpersonal relationships, may be appropriately partnered with direct personal interventions.

Conclusion

Investigating ethical wrongdoing by seniors reveals challenges for emerging technologies. By attending to the relational features that both enable and show promise in correcting moral errors among the elderly, it becomes apparent that trusting relationships provide both hope for rectification and reason to address that very wrongdoing. Engagement with the detailed accounts of filial piety available in the scholarship on Confucian ethics provides resources for thinking about how to preserve relationships with the elderly without subjecting ourselves to harm or licensing their misbehavior. What emerges is an injunction to take this wrongdoing seriously, while at the same time preserving affective bonds. While patients' comfort and safety remain important, and concerns about autonomy certainly have their place in design and use of geriatric care technologies, expanding ethical frameworks to capture the relational, moral-psychological, and interpersonal elements of patients' lives. Doing so helps us avoid falling into false dichotomies about tensions *between* caregiver and patient wellbeing, by reframing geriatric patient care in terms of promoting the flourishing of relationships that matter to both caregivers and senior patients.

Against framings of patient wellbeing that conceive of patient wellbeing as logically independent of and potentially in tension with moral behavior, and good patient care as a burden for caregivers rather than a source of value and meaning, the rich discussion of filial piety found among Confucian philosophers identifies both new areas of concern and new potential in the development of caregiving technologies, ones which see these goals as complementary.

Works Cited

Ames, R. (2011). *Confucian Role Ethics: A vocabulary.* Chinese University Press.

Burema, D. (2021). A critical analysis of the representations of older adults in the field of human–robot interaction. *AI & Society*, 1-11.

Cipriani, G., Vedovello, M., Nuti, A. and Di Fiorino, M. (2011), Aggressive behavior in patients with dementia: Correlates and management. Geriatrics & Gerontology International, 11: 408-413. https://doi.org/10.1111/j.1447-0594.2011.00730.x

Confucius. *Analects: with selections from traditional commentaries*. Trans. Edward Slingerland. Hackett Publishing, 2003.

Dennis, M.J., & Clancy, R.F. (2022). Intercultural ethics for digital well-being: Identifying problems and exploring solutions. *Digital Society* 1(7). https://doi.org/10.1007/s44206-022-00006-2

Dobrosovestnova, A., Hannibal, G., & Reinboth, T. (2021). Service robots for affective labor: a sociology of labor perspective. *AI & Society*, 1-13.

Draper, H., & Sorell, T. (2014, October). Using robots to modify demanding or impolite behavior of older people. In *International Conference on Social Robotics* (pp. 125-134). Springer, Cham.

Dreyfuss, Emily. (2018, December). The terrible joy of yelling at Alexa. *Wired Magazine*. https://www.wired.com/story/amazon-echo-alexa-yelling/ Accessed September 29, 2021.

Gerino, E., Caldarera, A. M., Curti, L., Brustia, P., & Rollè, L. (2018). Intimate Partner Violence in the Golden Age: Systematic Review of Risk and Protective Factors. *Frontiers in psychology*, *9*, 1595. https://doi.org/10.3389/fpsyg.2018.01595

Hempel, Jessi. (2015, October). Siri and Cortana sound like ladies because of sexism. *Wired Magazine*. https://www.wired.com/2015/10/why-siri-cortana-voice-interfaces-sound-female-sexism/ Accessed September 29, 2021.

Herr, R. S. (2003). Is Confucianism compatible with care ethics? A critique. *Philosophy East and West*, 53(4), 471-489.

Huang, Y. (2013). *Confucius: A guide for the perplexed*. New York: Bloomsbury.

Huecker MR, King KC, Jordan GA, et al. Domestic Violence. [Updated 2021 Feb 17]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK499891/

Ivanhoe, P. J. (1991). Character Consequentialism: an Early Confucian Contribution to Contemporary Ethical Theory. *The Journal of Religious Ethics*, 19(1), 55–70. http://www.jstor.org/stable/40015116

Lambert, A. (2016). 7 Confucian Ethics and Care: An Amicable Split?. In *Feminist Encounters with Confucius,* eds. Matthew Foust and Sor-Hoon Tan, . Leiden, Netherlands: Brill. pp. 171-197

Leggett, A., Polenick, C. A., Maust, D. T., & Kales, H. C. (2018). "What Hath Night to Do with Sleep?": The caregiving context and dementia caregivers' nighttime awakenings. *Clinical gerontologist*, 41(2), 158-166.

Li, C. (1994). The Confucian Concept of Jen and the Feminist Ethics of Care: A Comparative Study. *Hypatia*, *9*(1), 70–89. http://www.jstor.org/stable/3810437

Mencius. (2004). *Mencius*. Trans. D. C. Lau. Penguin UK.

Neville, Robert Cummings. (2017). *The Good is One, Its Manifestations Many: Confucian Essays on Metaphysics, Morals, Rituals, Institutions, and Genders*. SUNY Press.

Olberding, A. (2019). *The Wrong of Rudeness: Learning Modern Civility from Ancient Chinese Philosophy*. Oxford University Press.

---. (2015). It's not them, it's you: A case study concerning the exclusion of non-Western philosophy. *Comparative Philosophy*, *6*(2), 14-34.

Pang-White, A.A. (2011), Caring in Confucian philosophy. *Philosophy Compass*, 6: 374-384. https://doi.org/10.1111/j.1747-9991.2011.00405.x

Parks, J. A. (2010). Lifting the burden of Women's care work: should robots replace the "human touch"?. *Hypatia*, 25(1), 100-120.

Pew Research Center. (2019) "Generation Z Looks a Lot Like Millennials on Key Social and Political Issues"

Rosemont, H. (2018). Classical Confucian and contemporary feminist perspectives on the self: Some parallels and their implications. In *Culture and Self,* ed. Douglas Allen. New York: Routledge, pp. 63-82.

Schreier, J., Ford, C., Niederhoffer, G., Bisbee, S., Bisbee, J. K., Acord, L., Rifkin, D., ... Sony Pictures Home Entertainment (Firm),. (2013). *Robot & Frank*.

Shibata, Takanori. (2012). Therapeutic Seal Robot as Biofeedback Medical Device: Qualitative and Quantitative Evaluations of Robot Therapy in Dementia Care. *Proceedings of the IEEE*, 100(8), 2527-2538.

Wong, D. B. (2009). *Natural moralities: A defense of pluralistic relativism*. Oxford University Press.

--- (2021). "Chinese Ethics", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL =

<https://plato.stanford.edu/archives/sum2021/entries/ethics-chinese/>. Accessed May 29, 2022.

West, M., Kraut, R., & Ei Chew, H. (2019). UNESCO. I'd blush if I could: closing gender divides in digital skills through education. https://unesdoc.unesco.org/ark:/48223/pf0000367416.locale=en Accessed September 29, 2021.

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.