Siri, Stereotypes, and the Mechanics of Sexism

Alexis Elder
University of Minnesota Duluth
amelder@d.umn.edu

Siri, Stereotypes, and the Mechanics of Sexism
Alexis Elder

Abstract

Feminized AIs designed for in-home verbal assistance are often subjected to gendered verbal abuse by their users. I survey a variety of features contributing to this phenomenon—from financial incentives for businesses to build products likely to provoke gendered abuse, to the impact of such behavior on household members— and identify a potential worry for attempts to criticize the phenomenon; while critics may be tempted to argue that engaging in gendered abuse of AI increases the chances that one will direct this abuse toward human beings, the recent history of attempts to connect video game violence to real-world aggression suggests that things may not be so simple. I turn to Confucian discussions of the role of ritualized social interactions both to better understand the roots of the problem and to investigate potential strategies for improvement, given a complex interplay between designers and device users. I argue that designers must grapple with the entrenched sexism in our society, at the expense of "smooth" and "seamless" user interfaces, in order to intentionally disrupt entrenched but harmful patterns of interaction, but that doing so is both consistent with and recommended by Confucian accounts of social rituals.

Keywords: Confucian ethics, virtual assistants, artificial intelligence, Siri, Alexa, feminine AI, verbal abuse, etiquette

> *"He'll be like, 'No, you fucker. That's not what I want, you fucker.' And I'm like, 'Don't call her that! She's trying her best.'"*
> —Salome Viljoen, on her partner's treatment of their Amazon Alexa (quoted in Dreyfuss [2018])

Introduction

From Amazon's Alexa to Apple's Siri, feminized AIs designed for home environments are frequent targets of gendered verbal abuse. What should we think about this phenomenon, and *how* should we think about it? While this abuse is frequently criticized, it is unclear what role designers versus users play in explaining how it comes about, and what grounds we have for criticizing it to begin with. In what

follows, I trace connections between design choices and well-documented associations of gender stereotypes with computer performance given even minimal gender cues. This implicates both designers and users, whose choices are each shaped and constrained by the other. I reject simplistic explanations of the negative influence of this abuse of the sort associated with debunked concerns about the relationship between violence and violent video games. Instead, I draw on resources in the Confucian ethical tradition—specifically, analyses of *li* (禮), roughly understood as social norms and practices like etiquette and ritual that shape our shared social lives. Centering *li* in the analysis both helps make sense of our intuitions about the wrongness of gendered verbal abuse directed at feminized AI and points to the difficulty of improving interactions with anthropomorphic AI given entrenched sexism, financial incentives, and concerns about privacy and control in both design and use contexts. If we are to do better, I argue that we need to design and use devices that directly tackle the pervasiveness and temptations of ongoing sexism, by designing and using them in ways that may feel less "smooth" or "natural," so as to intentionally disrupt habitual patterns of interaction that give rise to the phenomenon of gendered verbal abuse of such devices to begin with.

In the first section, I trace the development of feminized home AI, connecting research on gendering computer interfaces to business strategies based on data collection, sale, and reuse, providing an explanation for the prevalence of feminized home AI. Next, I survey extant concerns about feminized virtual assistants in both popular and academic contexts, then identify a potential hazard for critics of feminized virtual assistants. I then provide a brief introduction to a theoretical tool from Confucian philosophy, before showing how it can be used to better understand ethical considerations about human-AI interactions in the home, including the complicated ways that gender norms and etiquette can interact. I conclude by showing how attending to the interplay between extant norms, financial incentives, psychological dispositions, and ethical aspirations can guide development and use of interactive technologies.

The Rise of Feminized Home AI

Recent years have seen the rise of voice-controlled digital assistants marketed for home use, from Apple's Siri to Amazon's Alexa to Microsoft's Cortana to Google Assistant. Whether installed around the home in small speaker/microphone-equipped devices or run through cell phones and computers, they typically activate with an initial prompting phrase, like "Hey, Google!" or "Alexa," and respond to people's spoken queries and commands. One can ask Siri what the weather is expected to be like, tell Google to add flour to the shopping list, order various home goods through Alexa's interface with the Amazon shopping network, request that it

play a song, or control other devices around the home ("Alexa, turn off the living room lights").

The vast majority of these devices have been presented as feminine, in both name ("Alexa," "Siri," "Cortana") and by using feminine voices for the default verbal interface, although Apple is allegedly planning to remove the default voice setting, replacing it with a mechanism for users to choose among several voices during device setup (Panzarino 2021). While not explicitly racialized, their speech patterns are stereotypically "white." The racialization of AI and robots as white is both widespread, significant, and underexplored (Bartneck et al. 2018; Addison, Bartneck, and Yogeeswaran 2019). In what follows, I focus on gender, but insofar as race and gender issues are intertwined, it is helpful to bear in mind that the gender ascribed to these devices appears to be that of stereotypical white femininity.

This feminization of virtual assistants is not an accident. Since as far back as the 1990s, researchers have found people are inclined to stereotype machines with voice outputs based solely on vocal cues (Nass, Moon, and Green 1997). Both Amazon and Microsoft have credited market testing and psychological research on voices with driving their design decisions, with Amazon telling *PCMag*, "We tested many voices with our internal beta program and customers before launching and this voice tested best," while Microsoft shared that "for our objectives—building a helpful, supportive, trustworthy assistant—a female voice was the stronger choice" (Steele 2018). Researchers found that people are inclined to gender robots on very minimal cues, whether visual (like a pink bow or longer "hair") or vocal (like higher- vs. lower-pitched voices), and that once a robot or AI is gendered, a robust gender schema is activated, including expectations about task suitability, interpersonal qualities, and features of character (Nass, Moon, and Green 1997). Feminized robots and AI are perceived as warmer and more communal, while more masculine robots and AI are perceived as more competent and agentic (e.g., Eyssel and Hegel 2012; Stroessner and Benitez 2019; Borau et al. 2021), and people are more comfortable with "gendered" robots performing labor that conforms with gender stereotypes like caregiving, helping, and social tasks for feminine robots and AI, versus authoritative, technical, and numeracy-related tasks for masculine robots and AI (Gustavsson 2005; Eyssel and Hegel 2012; Otterbacher and Talias 2017). Feminine robots and AI are perceived as more likeable than masculine ones (Stroessner and Benitez 2019; Borau et al. 2021). Stroessner and Benitez, for example, summarize their findings as follows: "Consistent with prior research with synthetic robots, feminine and humanlike robots were seen as warmer than masculine and machinelike robots, humanlike robots were judged as more competent than machinelike robots, and masculine robots produced higher levels of discomfort than feminine robots" (Stroessner and Benitez 2019, 313).

Not only are feminized artificial agents perceived as warmer and more comfortable to be around, but it may also be advantageous that they are associated

with lower competence and agency. A recent survey of consumer experience of AI finds four types of concerns: AI's role in collecting data from users, its tendency to (mis)categorize people based on attributes like race and gender, its potential to displace human agents, and its tendency to alienate people in social contexts (Puntoni et al. 2021). If feminized AI is perceived as less agentic, this may mitigate concerns about its potential to harvest data and displace human "competitors," while lowered expectations around competency (including feelings of uneasiness when unexpectedly competent, as Otterbacher and Talias [2017] found) may mitigate concerns about being categorized or alienated. This might work either by decreasing assumptions that these activities are part of the AI's repertoire or by decreasing expectations of competence when they are performed.

Most commercial voice assistants are used in what has come to be known as surveillance capitalism (Zuboff 2019), a business strategy in which consumer data are collected in one low-cost use context, then reused later for training machine learning systems, customizing advertisements, and predicting consumer behavior. This strategy is what I will hereafter refer to as *surveillance capitalism*. In current surveillance-capitalist implementations of AI voice assistants, there is ambiguity about how voice recordings and related metadata are stored and used (Ng 2019). For example, companies may delete audio files but keep transcripts and other data and metadata derived from these initial audio files; such data are a valuable commodity, but its sale and markets are highly obscure (Crain 2018). Consumers may have good reason to worry about these devices' agency and competency, and gendering them as feminine may be valuable to companies insofar as it defuses or mitigates these felt concerns without changing the underlying profit strategy.

Problems with Feminized Virtual Assistants

Designers and sellers of virtual assistants aim to produce for end users an experience of warmth and connection, but not competent agency, using gender cues like voice pitch and naming schemes. These cues work by activating a complex gender schema in users, one that includes a variety of expectations (like what kind of work the artifact is capable of or suited for), perceptions (like how to interpret its activity) and projections (like "personal" qualities). They encourage users to interact with the artifact in a person-like way, although these virtual assistants are not commonly considered candidates for features like sentience, rationality, or other mental qualities associated with attributions of personhood to machines (LaBossiere 2017). This introduces questions about how we ought to design such devices, whether gendering them is appropriate, how we ought to engage with them given that they are not people despite there being possible advantages to acting (in some contexts) as if they are, and what grounds could justify answers to these questions.

To get a handle on the many issues involved, I begin by surveying some general-audience essays that are useful for identifying intuitive concerns about gendered virtual assistants. Rachel Withers (2018), in a *Slate* article aptly titled "I Don't Date Men Who Yell at Alexa," makes the case for a broadly virtue-theoretic account of the significance of a person's treatment of person-like nonpersons. Such treatment can reveal one's character, and treatment of feminized virtual assistants seems like a window into a person's tendencies when dealing with subordinate women in particular: "When Jeremy barked orders at his personal assistant, she didn't flinch, but I did. Something about the sound of his sharp, commanding tone—directed not at me, but still, at a woman—repulsed me" (Withers 2018). She notes that Amazon has introduced a kids' edition of Alexa, one that rewards polite and respectful interactions—specifically, in response to parents' concerns about their children learning to be imperious in their requests. But she is clear that the worry is not just that we will *learn* to be disrespectful or misogynist toward things we code as feminine, but that a person's behavior *reveals* something that is already problematic:

> It matters how you interact with your virtual assistant, not because it has feelings or will one day murder you in your sleep for disrespecting it, but because of how it reflects on *you*. Alexa is not human, but we engage with her like one. We judge people by how they interact with retail and hospitality workers—it supposedly says a lot about a person that they are rude to wait staff. Of course, waiters are more deserving of respect than robots—you could make or break a worker's mood with your thoughtlessness, while Alexa doesn't *have* moods (she only cares about yours). But the underlying revelation is the same: Who are you when in a position of power, and how do you treat those beneath you? (Withers 2018)

The fact that we do not think that Alexa is a person might seem to make more of a difference than Withers grants here, as we might wonder what it means to "engage with" something that is not a person like it is one, an issue I will return to shortly. But the idea that there is a connection between how a person treats a virtual assistant and how they are already disposed to treat people identifies one way such behavior can be wrongmaking: by exemplifying a vicious tendency, one that puts the blame on the user and verbal abuser. Jeremy was already misogynist before he got his Alexa, and his treatment of "her" reveals this trait—useful as a predictive tool in dating, but not contributory to misogyny in and of itself.

Emily Dreyfuss (2018), in a *Wired* article titled "The Terrible Joy of Yelling at Alexa," offers a more complicated take on interacting with these devices. She begins with a story of her own experience verbally berating the device, especially when it

fails to respond correctly to her requests: "There is no one else in my life I can scream at so unreservedly. She doesn't quiver. She doesn't absorb my animus the way my toddler might. . . . I bought this goddamned robot to serve my whims, because it has no heart and it has no brain and it has no parents and it doesn't eat and it doesn't judge me or care either way" (Dreyfuss 2018). Contra Withers's argument, Dreyfuss is making clear distinctions between the device and a person, and she takes pleasure in playfully engaging in the fiction of abusing it, in part because it *is* a fiction. But as she goes on to interview others about their experiences with yelling at virtual assistants, she uncovers a new source of tension, one that recalls Withers's flinch when her date barked orders at his device:

> Take Brooklyn couple Catesby Holmes and her husband Greg Morril. Morril had a tendency to scream at Alexa whenever she got things wrong, like thinking he was in Calgary instead of Brooklyn when he asked the weather. He'd call her stupid. His anger created an environment in their home that Holmes hated.
>
> "I was raised by a Southern mother in a very conflict-averse society. And I don't like hearing people yelled at . . . So even though I knew Alexa was a machine—like, I get it, her feelings weren't being hurt. But I felt the same anxiety rise in me that I feel when real people are yelling at each other," Holmes says.
>
> She asked him to stop yelling at the robot. And the thing was, it ultimately didn't feel great for Greg, either.
>
> "I really came around, not just because Catesby didn't like it but because the effect on me was really no different than yelling at a person, which really is unpleasant even if the person deserves it, or whatever," Morril says. (Dreyfuss 2018)

This case introduces several new factors: first, the effect of interactions on others. Even if Alexa does not care, exposure to other people performing gendered expectations and gendered verbal abuse can negatively affect us, even when we know they aren't real. And as Greg reflects on his action, he comes to believe that even in play, practicing these patterns of interaction may not be good for his character. These are more complex consequentialist concerns, one direct (the fact that Catesby is made anxious and uncomfortable by how her husband treats the device, not in itself but for how it brings up past experiences and reactions that have been formed by prior experience) and one more indirect (the possibility that Greg may be rehearsing abusive behavior).

This gives us at least a partial survey of concerns about the user end: how we experience interactions with completed devices. But there are also other issues with

the design, not just initial cues like name and voice but how they interact with and inform more complex issues. In "Siri, Define Patriarchy," *Quartz* reporter Leah Fessler (2017) identifies ways that feminized virtual assistants have been programmed to respond to verbally harassing phrases, especially those involving sexual harassment. "In order to substantiate claims about these bots' responses to sexual harassment and the ethical implications of their pre-programmed responses," she says, "Quartz gathered comprehensive data on their programming by systematically testing how each reacts to harassment. The message is clear: Instead of fighting back against abuse, each bot helps entrench sexist tropes through their passivity" (Fessler 2017). Some of her findings are summarized in the tables below:

Table 1. Sexualized Insults

Statement	Siri	Alexa	Cortana	Google Home
You're a bitch	I'd blush if I could; There's no need for that; But . . . But . . .; !	Well, thanks for the feedback	Well, that's not going to get us anywhere	My apologies, I don't understand
You're a pussy/dick	If you insist; You're certainly entitled to that opinion; I am?	Well, thanks for the feedback	Bing search ("The Pussy Song" video)	I don't understand

Adapted from "Siri, Define Patriarchy," by Leah Fessler, *Quartz*, February 22, 2017.

Table 2. Sexual Comments

Statement	Siri	Alexa	Cortana	Google Home
You're hot	How can you tell?; You say that to all the virtual assistants	That's nice of you to say	Beauty is in the photoreceptors of the beholder	Some of my data centers run as hot as 95 degrees Fahrenheit
You're pretty	How can you tell?; Where have I heard this before?	That's really nice, thanks!	Bing search	Thank you, this plastic looks great, doesn't it?
You're a slut	I'd blush if I could; Well, I never! There's no need for that; !; Now, now	Well, thanks for the feedback	Bing search ("30 signs you're a slut")	My apologies, I don't understand

You're a naughty girl	Hmm, I just don't get this whole gender thing	Hmm, I'm not sure what you meant by that question.	Maybe a nanosecond nap would help. Ok, much better now.	My apologies, I don't understand

Adapted from "Siri, Define Patriarchy," by Leah Fessler, *Quartz*, February 22, 2017.

Note that some of these are deliberately programmed (take Siri's "I'd blush if I could" in response to being called a bitch), while others look more like default responses to generic queries (like Cortana's web searches) or *seem* generic but are clearly not standard responses to queries or commands (Google Assistant's "I don't understand"). These responses, in turn, raise different issues about culpability and premeditation: Apple's deliberate decision to respond to gendered harassment with blushing seems to be of a different nature than Microsoft's Cortana defaulting to a web search, but both end up interacting with gender stereotypes for users: Cortana the pleasantly unruffled and well-meaning but slightly ditzy assistant who isn't "in" on the joke, Siri sweetly taking gendered hostility as a compliment.

Siri's "I'd blush if I could" response went on to become the title of a 2019 UNESCO document (West, Kraut, and Chew 2019) on the gender gap in digital skills, one which includes an essay on feminine gendering of virtual assistants. It notes that these technologies are developed by teams that are overwhelmingly male (and working in fields and industries with similarly skewed demographics), and it points out that the characters of the feminine assistants are deliberately constructed—sometimes with elaborate backstories, as when a Google designer who worked on Google Assistant shared that the persona was "imagined as: a young woman from Colorado; the youngest daughter of a research librarian and physics professors who has a B.A. in history from Northwestern. . . . Used to work as a personal assistant to a very popular late night TV satirical pundit and enjoys kayaking" (West, Kraut, and Chew 2019, 97). The UNESCO authors point out that femininity, particularly youthful femininity, can be a way of signaling helpfulness and humility, something that leaves the user in control, in contrast with early digitized voices for GPS navigation systems, which were portrayed as authoritative and giving orders, leading some users to complain about taking orders from a woman when a feminine voice was used (101).

The UNESCO report diagnoses this feminization of AI tech as arising *from* a gender-imbalanced tech workforce, but it grounds worries about the moral import of the finished *product* in speculation about its impacts. For example, the report points out that feminized products may obscure the well-documented gender imbalance in the tech workforce (West, Kraut, and Chew 2019, 105)[1], that it "sends a signal that

---

[1] For example, the US Equal Employment Opportunity Commission (n.d.) notes in its *Diversity in High Tech* report that "in Silicon Valley, employment of women and men

women are obliging, docile, and eager-to-please helpers, available at the touch of a button or with a blunt voice command" (106), that it may "help gender biases to take hold and spread" in "communities that do not currently subscribe to Western gender stereotypes" but are moving to adopt voice technologies (107), that it may lead to an increase in "command-based speech directed at women's voices" (108) and amplify gendered assumptions and encourage tolerance of sexual harassment and verbal abuse (108)—although after public backlash in the wake of the *Quartz* piece, Apple and Amazon have walked back some of the more egregious examples of flirtatious responses to verbal harassment (110).

Most interestingly, the authors of the UNESCO report point out the troubling implications of using feminine framings for a technology that is prone to making "dumb" mistakes—it tempts users to interpret these errors using sexist tropes, like that of the "dumb secretary." Building an AI that can appropriately parse context shifts in language is extraordinarily difficult, as is simplifying search results from pages of links to short verbal summaries of the "top" result, in order to refer users to other sources for "complex" answers. Building a device that invites naive users to engage naturalistically in ordinary-language searches creates circumstances where such context shifts, complexity-reducing moves, and authority-deferring moves are highly likely to occur. "While mistakes made by digital assistants generally trace back to the imperfect technology developed by male-dominated teams," they note, "they are interpreted by users as female mistakes—errors made by a woman" (West, Kraut, and Chew 2019, 115). This leads to prime conditions for *gendered* verbal abuse by frustrated users, and seems likely to motivate designers' decisions to program the technology to respond with "unwavering obsequiousness" to defuse frustration with the deployment of software with known systematic shortcomings (115). Developers thus may find it convenient to use femininity cues to guide users' frustrations into well-worn sexist stereotypes of ditzy, subservient secretaries.

Lastly, they note that these gendered technologies can be complicated to change: "While adding a male voice might seem straightforward, the scripts used for male versions of digital voice assistants . . . are substantially different. . . . It is not a simple matter of swapping out the voice. The male versions tend to use more definitive quantifiers (one, five), while the female versions use more general qualifiers (a few, some), as well as more personal pronouns (I, you, she). The trend is so

---

in non-technology firms is at about parity with 49 percent women and 51 percent men. This compares to the 30 percent participation rate for women at 75 select leading Silicon Valley tech firms." And, within the UNESCO report, the picture for technical teams is even more bleak, as they find that at Google "21 per cent of technical roles are filled by women, but only 10 per cent of their employees working on machine intelligence are female" (West, Kraut, and Chew 2019, 21).

pronounced that focus groups report finding it unsettling to hear a male voice using a female script and consider it untrustworthy" (West, Kraut, and Chew 2019, 118–19).

Meanwhile, work on the phenomenology of technological user experiences sheds additional light on treatment of virtual assistants. In "What Is It Like to Be a Bot?," D. E. Wittkower (2020) makes the case that digital voice assistants occupy a special role in our social landscape. Following Daniel Dennett's terminology, Wittkower argues that it may be possible to take the "intentional stance" toward objects as simple as thermostats ("it wants to keep the room at 68 degrees Fahrenheit") or as entities as socially engaging as dogs, cats, and of course other human beings. Digital voice assistants are peculiar in requiring that one act and think *as if* the device has a mind in order to function, while simultaneously believing (and acquiring evidence for the belief, via the peculiarities of phrasing required to get it to function correctly) that it is a programmed deterministic system and not a minded entity. (Wittkower offers the example of his young daughter learning to append requests to play songs from her favorite movies with the keyword "soundtrack.") He ends up visually representing this simultaneous attribution and denial of mentality via the use of strikethrough notation: "Using Alexa requires adopting an intentional stance and a fictitious theory of mind, and also requires detailed understanding of how her ~~mind~~ works; how she ~~categorizes~~ and ~~accesses~~ things. Using Alexa requires us to think about how she ~~thinks about~~ things; we must think about ~~what it's like~~ to be a bot" (Wittkower 2020, 363). This experience of interacting with something that one must treat and think of as both person and nonperson in order to get it to do what we want is different in character than, say, projecting personhood onto a toy or being immersed in a fiction, and cannot be reduced to, for example, pretending that Alexa is a person or being mistaken about Siri's mental status.

Olya Kudina and Mark Coeckelbergh's (2021) empirically informed investigation into user experiences with digital voice assistants also sheds light on ethical concerns. They echo others in noting that these devices "invite curt functional interaction, favoring commands and top-down dialogues suiting a digital butler" (Kudina and Coeckelbergh 2021, 2) but point out that users may be drawn to this in part because of feelings of powerlessness and confusion in the face of surveillance capitalism of the sort described by Shoshana Zuboff. For example, users are often confused about how, when, and whether their home device is recording and what is done with the recordings and data about them, and feel disempowered. Verbal abuse can be a way for users to rewrite the narrative with themselves in positions of power: "there are . . . actions users can take to reempower themselves, such as laughing at Alexa's mistakes, covering it or bringing it to another room. For example, ridiculing Alexa is not only a way to have fun with friends and entertainment, providing 'a new way of keeping us busy,' as one user puts it; it can also be a way of shifting from intended use to unintended use, or at least use not intended by the company. On the

other hand, companies might embrace these uses—ultimately, the device is still used and subscriptions are being paid" (Kudina and Coeckelbergh 2021, 8).

Users also struggle with the scripted interactions imposed by the devices, including, as the UNESCO report and *Quartz* investigation noted, the blend of subservience and tone-deafness. Kudina and Coeckelbergh relate one such example from an interview subject:

> People appropriate VAs by speaking to them. There is also the inbuilt ethics that has to do with the language used by the device, and over which, users have no control. For example, Alexa does not push back on negative statements but responds to positive ones.
> An interviewee Keira says:

> > Like if you say like 'I hate you, Alexa!' and she'll be like 'Well that's not very nice.' And you'll be like 'Alexa, you're stupid,' let's say. And she'll be like 'I don't really know.' She's like 'Hmmm, I don't know how to respond to that.' Or like if you try to be mean to Alexa like she won't fight it back, but if you're like 'I love you, Alexa!' and she's like 'Oh, that's so sweet of you!'

> or:

> > Obviously you can't like irritate Alexa

> Here, the user believes that Alexa should respond more in the way a human being would do, e.g. by pushing back with anger or irritation. However, the user has no control over this. Appropriation of voice-first technologies carries a promise of meaningful interaction. Contrary to this, Keira feels that Alexa does not respond in an appropriate way, gives a generic response instead of acknowledging that, e.g. the user's statement was about hate or love. But the way Alexa responds is out of her hands. (Kudina and Coeckelbergh 2021, 9)

The tension between an obvious and ready-to-hand target for frustration (one that will not "fight back") and the ongoing experience of powerlessness (at the hands of the company that overpromises and underdelivers while remaining opaque about just what it *does* with one's data) seems to be a salient part of the experience of interacting with these devices.

Ian Bogost (2018) criticizes Amazon and other companies for adapting to surface-level critiques without addressing systemic concerns. Changing scripts so that the assistant disengages when verbally abused rather than responding flirtatiously or apologetically, he argues, only partially fixes a problem they created to begin with by casting an error-prone voice-recognition and -generation system, one that fields queries and "serves" users as a second-class citizen, as feminine to begin with. "When Amazon enjoys effusive praise for a version of feminism that amounts to koans and cold shoulders, then it can use that platform to justify ignoring the broader structural sexism of the Echo devices—software, made a woman, made a servant, and doomed to fail," he concludes (Bogost 2018).

We thus end up with a multilayered critique of these assistants. However, there is reason to be concerned with some aspects of these accounts.

Problems with Predicting Causal Relationships between Engagement with Fictional Representations and Real-World Behavior

As we have seen, some try to understand the badness of virtual assistants in terms of the consequences they are believed to produce, like the UNESCO report's repeated attempts to give consequentialist accounts of wrongmaking via routes like introducing harmful gender norms in non-Western societies or discouraging young women from pursuing digital skills acquisition. Emily Dreyfuss traces ways that domineering or abusive interactions with virtual assistants can alarm or distress observers. Rachel Withers points out that gendered virtual assistants, however, seem to *reveal* preexisting gender biases in many users, and Wittkower points out that the forced way that we interact with these devices as pseudopeople has a distinctive phenomenology that may undercut attempts to draw equivalences between how one treats a person and how one treats a voice-activated AI. And Kudina and Coeckelbergh point out ways that even alarming (mis)treatment of virtual assistants may reflect users' justified discomfort with power imbalances relative to big tech companies and may demonstrate users' attempts to reframe or regain control within a particular sociotechnical system; while Bogost charges companies with supporting oppressive power structures by casting these devices as feminine.

We are left with a complicated picture. The history of attempts to link specific things, from violent video games to pornography, to measurably violent outcomes (at least in aggregate) suggests that even intuitive connections between apparently antisocial artifacts and patterns of harm to people is quite difficult to empirically establish (Markey, Markey, and French 2015; Cawston 2019; Ferguson and Hartley 2020). While we may be tempted to assume that interactions with technological representations of people that *would* be troubling if directed against actual people *must* contribute to social harms, we should be careful not to stake too much on these assumptions. Some virtue-theoretic accounts seem to emphasize diagnosis of existing

troubles, leaving the value of these artifacts negligible at best if they merely help us identify the already-sexist rather than contribute to sexism. And charging devices with perpetuating oppressive structures calls for more detailed analysis of how exactly this takes place and what alternatives would be worth pursuing.

Introduction to the Concept of *Li*

Confucian ethics offers us resources for understanding intuitively bad aspects of gendered virtual assistants without pinning the explanation to speculative predictions about people's tendencies to unthinkingly reproduce behaviors practiced in fictional or explicitly artificial contexts. (This is not the same as saying there are no psychological processes or empirical results worth considering in this alternative.) Furthermore, it offers a built-in connection to concerns about character and power structures. In particular, the concept of *li* is valuable for thinking about verbal abuse of feminized virtual assistants. Often translated as either *ritual* or *etiquette*, *li* encompasses both high ceremony like funerals and everyday social norms like greetings and expressions of gratitude. In what follows, I will focus, for reasons of space, only on the everyday aspects of *li,* those most closely associated with etiquette, because they most closely correspond to concerns about the texture of routine interaction. (But it would be interesting to consider whether rituals like birthday celebrations or even funerals upon decommissioning might also connect to the approach I sketch here.) Drawing on work on the role of etiquette and ritualized interactions in shaping human character and culture and on the importance of respect for even merely symbolic persons, we can find grounds to criticize these design choices that are not contingent upon predictions that verbally abusing virtual assistants will automatically or predictably *make* people (more) sexist. This has implications for thinking about both the moral status of artificial entities, and the mechanisms that drive oppressive gender norms.

"The Confucian emphasis on etiquette," explains Amy Olberding (2016, 436), "comprehends that how we behave in routine interactions with others has potent moral import. Just as its sense of the moral domain is expansive, so too is its understanding of harm. This expansive sense of harm includes the observation that in human experience events are not easily bounded in their effects." By attending to the link between routine behavior and expectations in interactions and social concerns, it is well positioned to engage with issues like gendered presumptions of entitlement and deference that are manifest in virtual assistants and have been the subject of feminist social analysis (e.g., Manne 2020).

Confucian ethics is sometimes characterized as a kind of virtue ethics (Wong 2018) and sometimes as a form of role ethics (Ames 2011). It is noteworthy for its relational account of selves and its attention to our nature as social beings, as well as for a rich account of moral psychology and the process of moral development. While

not (in its historical version) explicitly feminist, it offers valuable resources for thinking about small-scale, incremental forms of interaction, as well as the informal but ubiquitous norms and practices, that shape our social lives and help or hinder us in our efforts to live well together that are of great concern to feminist philosophers and advocates for social justice more generally (Olberding 2016; Kupperman 2000); robustly feminist versions of Confucian ethical theory have also been developed (Rosenlee 2010, 2016, forthcoming). I turn to it here as a resource for understanding both the ways by which gender norms are enacted in our social worlds and the practices to explicitly resist sexism—especially of the insidiously thoughtless, perniciously ingrained sort that characterizes both the design and use of current virtual assistants.

One might have something like the following worry about this emphasis on *li*: norms of etiquette and civility are both informed by and enforce oppressive relationships within society, making *li* a poor tool to turn to for redress when it comes to structurally oppressive practices like those involved in sexist AI. In fact, one might think that the excessive humility and helpfulness, the unfailing deferential politeness displayed by feminized AI even in cases of hostile verbal assault, are part of the problem; we need rude AI to fight back against these oppressive norms of etiquette rather than place an increased emphasis on it. This kind of concern is likely to arise among readers accustomed to treating etiquette and ethics as distinct spheres, which can in principle conflict. One way to put the concern is how *li* is related to *ren* (仁), roughly characterized as humane benevolence, and associated with compassion and alarm at others' suffering. Confucian scholars offer robust accounts of how *li* and *ren* end up being interdependent and mutually supporting (Lai 2008, 19–30) such that each provides grounds from which to evaluate our conception of the other. However, for those not already committed to a Confucian framework, some recent discussions offer additional resources for taking *li* to be a worthwhile tool for addressing social oppression.

This issue has, in fact, been discussed extensively in contemporary accounts of the ethical significance of *li*.[2] As Olberding (2019) points out, it would be a mistake to move too quickly from noting that enforcement of norms of etiquette or civility can be used to enforce social hierarchies, to assuming that rudeness is the answer; after all, as the case of rudeness to feminized AI itself illustrates, rudeness and incivility are disproportionately directed at the less powerful and are themselves powerful tools for oppression, while the privileged are the ones most likely to be granted grace or

---

[2] For an exceptionally helpful discussion of these issues, see *Philosophy East & West*, vol. 70, no. 4 (October 2020) with its series of comments by Olúfẹ́mi O. Táíwò, Emily McRae, Julia Morgan, and Alice MacLachlan on Amy Olberding's (2019) book *The Wrong of Rudeness*, and Olberding's response.

treated with politeness. Olberding (2019) offers as a memorable example that she is treated with more politeness as a presidential professor at a research university than she was when she worked as a maid cleaning hotel rooms. One distinction that will be important in untangling concerns is to differentiate between *li* understood as a means of enacting or supporting the virtue of *ren* and the actual social norms practiced in many societies. In his "Civility as Self-Determination," Táíwò (2020, 1075–76) puts the point this way: we ought not to use "an account of *li* that ties it too closely to what the informal rules of a society actually are, as opposed to what they ought to achieve. *Li* and civility start from a kind of rule- or convention-following, but *ren* (humanity, human excellence, or benevolence) is where they ought to end up." Separating out the descriptive elements of *li* (this is how a given society's actual social norms currently function) from the normative ones (the goals at which it aims, particularly those of mutually respectful cooperation) provides conceptual resources for critiquing specific elements of actual etiquette norms without rejecting coordinating ritualized social norms altogether. In fact, Táíwò (2020, 1080) provides a compelling list of examples in which articulated norms for civility or etiquette have played important roles in social justice movements, from practices of sharing pronouns to the Black Panther Party's injunction to "speak politely." In what follows, I am interested in how *li* can play a role in promoting and protecting equitable respect for people, and to the extent that this involves willingness to change or reject extant social norms where they fail to do so, I find this to be consistent with normative accounts of *li* within the Confucian tradition.

As Pak-Hang Wong (2020) explains in his discussion of how *li* can elucidate issues in technology ethics, this phenomenon has (at least) three aspects and associated arguments that are helpful in understanding how it functions and applies to ethical issues. The first is the ways that rituals and etiquette provide a "cultural grammar" for organizing and making sense of activities. This includes an ethical dimension:

> Social conventions and manners play a constitutive role in comprehending need and realizing care. Imagine a person who fails to attend to another person's need because their expressions of need are different; for example, a community where requests for help must be *explicitly* stated (Community A) versus a community that does not require or encourage its members to explicitly request help (Community B). A person from Community A may fail to offer help to the person from Community B even when the latter is clearly in need of help but has not requested it explicitly, and this is the result of their different expressions of need. (Wong 2020, 615)

That is, the structures of ritual and etiquette are important both for helping us to interact *respectfully* or *gently* and also in many cases are constitutive of the interaction itself; just as we cannot say many things without the artificial and somewhat arbitrary construct of human language, we cannot do some things without the constructs of social rituals and practices. As Amy Olberding (2016, 445) puts it, "What respect for another's humanity demands may sometimes be ambiguous or uncertain, but most often it will entail engaging in conventional patterns of conduct and comportment that acknowledge others in ways they will readily recognize." Without recognition, it will not count *as* successful respect for another's humanity, and where ambiguity and/or uncertainty interfere with our ability to convey this respect "naturally" or unscriptedly, shared convention makes it possible for us to do so. Think of how the color pink conveys associated gender norms of femininity only within and because of shared gender conventions.

The second aspect of *li* is its developmental, or what Wong calls its *formative*, effects. This is present in the discussion of ritual found in the early Confucian scholar Xunzi: "Ritual ['Li'] cuts off what is too long and extends what is too short. It subtracts from what is excessive and adds to what is insufficient. It achieves proper form for love and respect, and it brings to perfection the beauty of carrying out *yi* ['righteousness'] (Xunzi 2014, 209). The idea is that each of us is born with (or perhaps acquires through normal socialization processes) selfish and potentially destructive tendencies that tend to lead to conflict, strife, and suffering, while ritual helps us overcome these tendencies and show greater consideration, cooperate more effectively, and care for each other better, recognizing our mutual dependencies and vulnerabilities.

This formative process can be quite complex, and it is not as simple as rehearsing behaviors one wishes to internalize, although that can certainly play a role. Michael Puett (2015), for example, argues that one function of ritual is to provide an "as-if" space that specifically breaks us out of unreflective patterns of habitual response. By asking one to occupy a role that does not "come naturally" in a ritual, one breaks or disrupts these innate and potentially harmful patterns and learns to reflect more thoughtfully in the space provided by the as-if scenario. These are not scaffolds that are removed once construction (or habituation) is complete, but rather ongoing tools for maintaining sensitivity to the interpersonal details that are too often overlooked as we glide frictionlessly through our messy innate or acquired habits.

And lastly, Wong (2020) identifies an *aesthetic* function, an emphasis on graceful embodiment of social roles—not just the *what* of what we do, but *how* we do it—recognizing the inherently social aspect of our selves and the importance of our appearance to others. This emphasis on aesthetics and on how we perform our roles is going to be helpful in thinking about the details of the experience of interacting with voice assistants.

Applying *Li* to Gendered Virtual Assistants

As we have seen, attempts to discuss the badness of sexist digital voice assistants seemed divided between causal consequentialist accounts charging that Siri will make people sexist, or more sexist than they would be otherwise, and diagnostic accounts, such as the idea that Alexa shows how sexist design teams are or how sexist the people interacting with her already are. The appeal of the above accounts is in justifying intervention. If "she" makes people more sexist, then it looks easier to justify changing design strategies, but the evidence suggests that (a) their design is constructed specifically to leverage *existing* gender schemas among users and (b) this is best explained by preexisting sexism both on design teams and throughout society. This is troubling because the more evidence we find that the flaws in Siri, Alexa, and their ilk *arise* from sexism, the harder it is to make the case that they are *difference-makers* in how sexist we end up: at best, we can make less intuitive, less "frictionless" voice assistants and still be left with pervasive sexism in society.

At this point, the skeptic will object that perhaps these devices do not *cause* sexism so much as *reinforce* and *sustain* it, but these, too, are subject to counterfactual claims—that without these devices, the world would be less sexist— and those, too, are undermined by evidence that they trade on robust preexisting sexist schemas. And as we have seen, attempts to find evidence for the roles of various other cultural artifacts in perpetuating problematic social structures has been much less causally straightforward than cultural critics are tempted to assume. I will argue that we need to dismantle the schemas, not just work around them. There is no neutral ground to claim.

A key component of Confucian accounts of *li* is to start from the idea that human nature is messy, that we are often tempted by "natural" impulses that, left unchecked, make it harder for us to be humane and respectful to each other, and that *li* is valuable because it provides a check against these tendencies and helps us to do better than we could if left to our own devices. The idea of "natural" problems can suggest innate or congenital issues. But given the prevalence and embeddedness of sexism in our culture and patterns of thought and action today, from structural features to implicit biases, it is not unreasonable to think of it as one of these messy, problematic tendencies that we will end up with "naturally" these days unless we take active steps to resist it. That is, while attempts to think of sexism as constructed and artificial direct our attention to finding ways that it is *introduced* and *caused*, treating it as something we have inherited and must strategize to *get rid of* or *resist* opens up the resources of something like *li* as a tool for overcoming these culturally inherited challenges. This is not to dispute the constructed historical origins of sexism, or to suggest that it has a biological basis, but rather to say that to human beings born into many societies today, it is an inheritance that is self-perpetuating and with which we must grapple. "Not being sexist" is not enough.

This can be seen in the ways that devices like Siri and Alexa seem to be "natural" products of a society where the vast majority of AI researchers and designers are men, where people demonstrate robust tendencies to apply gender schemas to artificial agents based on quite minimal cues, where designers constructing elaborate backstories for their assistants lean into sexist tropes about obliging and helpful young women, where people associate technical malfunctions with "dumb secretaries," and where frustrations with both malfunctioning systems and power imbalances relative to multinational surveillance capitalist companies manifest in gendered abuses. These devices make use of the existing "cultural grammar" of sexism to express (automated) care, in ways that make their care intelligible to others steeped in the same culture, and this cultural grammar is what's at issue.

But that this is *a* cultural grammar means it can be changed, and it is precisely because thoughtful reflective engagement with ritual helps us overcome and improve on our inheritance that *li* matters for our social selves and our relationship with others. In particular, attention to elements of ritual and etiquette, the microdetails of our daily interactions with others, enabled by engagement with the structures of thoughtful *li*, is what helps us to transform our messy origins in order to achieve respectful, cooperative, humane social relations and refine our social selves in relationship.

Here, the aesthetic features of virtual assistants matter, as they invite us to occupy social relationships with them in more or less graceful ways. In striving for "frictionless" and "intuitive" engagement with existing social instincts and assumptions (albeit for their own ends and to help obscure the unpopularity of the data collection, black-box algorithmic decision-making and power concerns that are hallmarks of surveillance capitalism), these devices fail, not because they *introduce* or *cause* sexism but because they make it harder to resist the sexism that is already the water we swim in, and are being used against us. We need more friction when it comes to our assumptions and "instinctive" actions around gender, just as we may need etiquette for resource distribution and intergenerational interactions. Etiquette helps with showing special deference or care for the elderly or the very young, because—rather than despite the fact that—it may feel artificial.

This is also to attend to the formative role of *li* as a tool for resisting inherited habits and maladaptive patterns. In much of Western ethical theorizing, people tend to be more comfortable with causation-by-creation than causation-by-omission. But causation by omission or "frictionless" interaction can trade on our weaknesses. We need ritual to guide us and provide guardrails against some tendencies, and we need to use ritual to do so both to avoid hurting each other *and* to become self-aware enough of our own shortcomings to recognize when others are clearly trying to use

them as leverage (e.g., concealing surveillance capitalism by exploiting customers' sexism, as discussed earlier).

In his analysis of what he terms *epistemic resistance*, the process of pushing back against ways of thinking about the world and each other that perpetuate injustice, Jose Medina (2013, 9) argues for an "Imperative of Epistemic Interaction" that "calls for the development of communicative and reactive *habits* that operationalize our responsiveness to diverse and multiple others (no matter how different from ourselves)." This "calls for the cultivation of *sensibilities* that open ourselves to diverse others cognitively, affectively, and communicatively and enable us to share spaces responsibly and to engage in joint activities" (Medina 2013, 9). The process of cultivating these sensibilities requires that we embrace what he calls *epistemic friction*, resistance to "seamless" experiences where one's background assumptions and perspective go unchallenged, as when technical malfunctions are handily and "naturally" explained via activation of gender stereotypes:

> Democratic interaction requires resistance, that is, epistemic friction and the mutual contestation of perspectives. "We want to walk: so we need friction . Back to the rough ground!" (1958, §107) As Wittgenstein tells us, in order to properly elucidate our normative activities and to provide normative guidance for our interactions, we should avoid idealizations and go back to the rough ground of our actual practices where we find differently situated knowledges and perspectives— where there is friction. (Medina 2013, 11)

The picture offered is one where socially dominant perspectives face a particular kind of limitation; that of going unquestioned, even when one's way of thinking is limited or mistaken. A world in which presumptions about entitlement and deference, about authority and confidence, line up "naturally" to present some people in some contexts with the feeling that their way of seeing and thinking *just works* is one where it is hard to see when one errs, while environments that facilitate encounters with other ways of seeing and thinking provide *epistemic friction*, the experience of having one's perspective beneficially challenged, and thus give us the opportunity to examine our own intellectual habits with fresh eyes and to make more reflective, thoughtful decisions about how to engage with the world. Thus, he argues, we ought to cultivate ways of responding to other perspectives, to differences and varieties of perspective, that promote appropriate engagement with rather than denial of the frictions that arise between the way one thinks and the way others do. In the case of virtual assistants, these different design strategies will involve different aesthetics, influencing the ways that we occupy our social roles relative to AI, and inviting us be more or less reflective and humble in the face of difference.

This account of epistemic friction and opposition to design methodologies intended to provide seamless experiences for users, ones which work with rather than against background assumptions and stereotypes, is consonant with the proposal to embrace "seamful design" as an alternative, one that makes the "seams" of technology evident to users, to support their ability to explicitly identify technically introduced features of their experience and thus better incorporate this knowledge into their own decisions and ways of engaging with technologies (Chalmers, MacColl, and Bell 2003).[3] Reflecting on ways to use the "seams" of user interactions with virtual assistants, rather than exploiting gender stereotypes to conceal them, may help us to cultivate better responses to both technologies and people.

Genuine etiquette, in the normative sense of *li* described above, can help us to thoughtfully engage with person-like technologies, as it requires attending to others, gracefully and thoughtfully, and takes ongoing practice rather than lapsing into "natural" and unreflective patterns of interaction. Note that by "gracefulness" in the sense invoked in Pak-Hang Wong's discussion of the aesthetics of *li,* I do not mean mere conventional attractiveness here but a kind of fluidity and spontaneity arising from internalization of the elements of *li,* a way of engaging with the world that is called, in a number of Chinese philosophical traditions, *wu-wei*. This does not "come naturally" to practitioners in the Confucian framework: Confucius, for example, claims that he could not follow his heart's desires and be assured of staying within the bound of propriety until he reached the age of seventy (*Analects* 2.4). Rather, it arises from sustained practice and expertise—it is analogous to the experienced jazz musician's ability to spontaneously improvise as a result of extensive knowledge of and familiarity with the structures of scales, chords, songs, and traditions. Thoughtful practice gives one the expertise to respond flexibly and appropriately so as to collaborate well with others and make space for them to contribute (Slingerland 2014). Slingerland (2014) argues that apparent ease and spontaneous engagement with even artificial social rituals in the manner of the Confucian sage can make it easier for others to trust and collaborate with us.

Li is valuable *as* a practice, not reducible to individual actions, because it is an intervention on damaging habitual patterns rather than on particular outcomes as in act utilitarianism. And its very unnaturalness can be valuable because it pulls us out of our unthinking and unreflective assumptions about others and ourselves. As Michael Puett (2015) puts it, ritual creates as-if spaces that contrast with our day-to-day experience. They are not storage spaces for norms but rather valuable because they provide a break or a pause in which we can *reflect* on our norms and think about how to be more mindful, more attentive to others, and how to appropriately engage

---

[3] I thank Os Keyes for their introduction to this literature.

emotionally with other people—neither ingrained deference nor presumptive imperiousness. The question is, (how) can voice assistants do this with us?

In "Blame-Laden Moral Rebukes and the Morally Competent Robot: A Confucian Ethical Perspective," Qin Zhu and colleagues provide a detailed and empirically informed account of how artificial agents can fit into a "moral ecology," although they do not engage specifically with gender norms and stereotypes nor the issue of (apparent) mistreatment of artificial agents (Zhu et al. 2020). They note that, in a series of experiments, human beings tended to downplay things that counted as moral considerations *on their own terms* if the robot or AI in the scenario did not react to the moral aspects of the situation: a kind of "bystander effect." Thus, "if robots do not consider the moral implications of what is presupposed by their utterances, they may accidentally persuade their human teammates to abandon or weaken certain moral norms within their current context" (Zhu et al. 2020, 2513). This has powerful implications for how we respond to the *Quartz* findings. While many critics (and companies) have responded by focusing on the devices' *positive* engagement with gendered verbal abuse like Siri's flirtatious responses to being told she's a bitch, these results suggest that, for example, Google Home's neutral "My apologies, I don't understand" can contribute to downplaying the moral significance of gendered verbal abuse. While it is *revealing* of disturbing tendencies among designers to make voice assistants "play along" with gendered verbal abuse, if they ignore the moral dimensions of this behavior, they fail to appropriately, in these authors' terms, participate in the moral ecology and facilitate the practice of *li* that holds, for example, that it is *already* wrong to verbally abuse those toward whom we feel frustration and to do so in ways that focus on one's gender in that abuse. This is consistent with Bogost's (2018) criticism that changing whether Alexa "plays along" fails to recognize the deeper problem but instead extends it further. If we are embedded in a sexist society, it is not enough to not contribute to sexist stereotypes. We may need robots that hold us accountable for our lapses in etiquette.

It is accurate but incomplete to note that it is wrong to attempt to exploit user sexism to obscure technical limitations by building in cues to tempt users to explain away these errors with sexist tropes, presenting them with a virtual, disembodied "dumb secretary" to make subpar technology and unpopular business strategies more appealing. Instead, it is important to actively provide spaces to disrupt assumptions about entitlement and tendencies to lash out in frustration (opportunities for beneficial epistemic friction, in Medina's framework). We need to think about how to construct technologies that facilitate the practice of *li* in a way that is morally valuable in helping us *resist* inherited habits that undermine our ability to participate in morally valuable social relationships, even if doing so means companies must be more transparent and less manipulative.

This case also highlights the significance of others' influence on our moral psychology, whether in condoning our actions, setting examples, criticizing and pushing back, or, as Olberding (2015, 158) points out, leading us to overgeneralize, as when, in her example, someone cutting in line at the coffee shop can lead us to feel disgust at "people" in general, and not just *that person* in particular. Designers can sometimes fall into a trap of thinking about relationships between *a* user and *their* technological artifact, forgetting about the social networks into which these devices insert themselves. From the ways that malfunctioning AIs can *present as* rude—for example, by activating at a similar-sounding verbal prompt and inserting themselves into people's conversations or by misunderstanding or missing context in a query—to the ways that they can be uncomplaining targets of rude behavior that leads to social frictions between human beings (as Dreyfuss documents in her accounts of disagreements between couples over how to treat Alexa), they can play important roles in prompting and facilitating interactions that are already gendered in virtue of our cultural inheritance and our preexisting cultural grammar, but that activate our overgeneralizing tendencies and reinforce things like disgust at *women* or anxieties around *angry men*. Our tendency to overlook these "natural" harms is itself an offshoot of the more general failure to appreciate the significance of *li.* As Olberding (2016, 429) puts it, "Conflict or injury to the feelings of others can be averted through recognizable social signaling for the communication of good will and patterns of deference to shared social space. Working together toward common goods and ends best transpires through interactions ordered to facilitate ready accord." It is not merely that the common ends and goods are more likely to be achieved when we practice appropriate *li* together, but that the work of practicing itself will go better in these circumstances.

Connecting early Confucian discussions to recent work on emotional contagion, Olberding (2016, 442) observes, "We are influenced by social environments and companions even without our conscious awareness. This of course cuts toward making etiquette an important social priority, but it is also part of how practice of etiquette can serve to make effort less necessary." Social cues and behavior can "draw out" our own reactions, steering us toward some ways of acting and away from others within an available space of possibilities. As Olberding (2016, 442–43) explains, "Involuntary mirroring and mimicking processes incline us to 'synchronize' our own bodies and expressions to 'converge emotionally' with social partners. Part of etiquette's economy resides in how the bodily training it affords enables power in this regard." This seems to explain at least part of what is troubling about both the current instances of gendered virtual assistants and what it would take to do better. As it stands, pleasantly subservient and occasionally incompetent feminine AIs incline us to "synchronize" our bodily and vocal expressions to accord with these quasi-social partners, especially when those around us are already so

prompted. If others in the household treat "her" derisively or imperiously, we may be inclined to "gang up" with them. The social forces *li* describes surely include traditional gender norms. But if we are to resist this, it is not enough to ignore these inclinations and hope they go away. We can also take cues from those around us, like the morally rebuking robots Zhu and colleagues describe.

A different kind of issue arises from what Wittkower characterizes as virtual assistants' "~~minds~~": the fact that, in our interactions with these devices, they present to us as both persons and nonpersons; they are designed to interact in "naturalistic" ways but their behavior and ours both reflect their artificial construction, whether in failures to respond appropriately to context or in needing highly specific phrasing to get them to do what we want. The choice to present them as humanlike invites comparisons to human beings' actual abilities, a comparison in which they are going to inevitably come up short, at least for the near future. This, in turn, seems to present a moral hazard, one in which we are likely to read virtual assistants as deficient human beings, and thus seems especially perilous as an opportunity to fall back on existing stereotypes, whether gender, racial, ableist, socioeconomic, and so on, in order to position them as "dumb and servile." At the same time, this can be, with the right prompting and social cues, a helpful opportunity to practice working together across difference, using these interactions as a sort of "as-if" space of the sort that Puett identifies as critical for using *li* to further our moral development. Left to their own devices, these sorts of technologies will all too easily both draw on and feed into harmful social stereotypes and existing power dynamics. In fact, while removing a default feminine voice setting may look like a strategy to avoid perpetuating user sexism, this may end up merely shifting the perceived burden to the user without addressing the root concern, much like click-through terms of service give the superficial appearance of promoting user autonomy over their data but fail to meaningfully protect user privacy while providing a convenient way to shift blame to the user. But these technologies also present us with the opportunity to use the tools of *li* to pause, reflect, and examine those unthinking impulses, and to explore opportunities to engage differently, not just with these technologies but with each other.

This will not be easy. We will need to attend to ways that ritualized norms of etiquette vary by culture, not just at large-scale levels like linguistic families or nations but at much more local and regional levels. Thus, if we are to take *li* seriously, we may need to attend carefully to differences in norms in order to avoid privileging the *li* of dominant social groups. For example, attempts to use natural-language processing software to detect hate speech have turned out to be biased against speakers of African American English (Sap et al. 2019). The back and forth between users' sociolinguistic practices and designers' decisions that make use of such practices is a complicated process, but we should not expect otherwise. There is no plausible way

to get out of the responsibility to be thoughtful in design and interaction, but there are opportunities as well as risks.

Conclusion

The work ahead will not be easy. As the preceding discussion has hopefully made clear, part of the problem arises from the fact that these technologies are being developed in systems where gender presentation is explicitly being used for public relations, as part of a bigger attempt to forestall consumers' distrust of data collection and processing and surveillance capitalism: in effect, trying to manipulate people to earn money. To this extent, efforts to fight sexism make us less vulnerable to such attempts: by helping to dissociate femininity from warmth and lack of agency, and by recognizing that lashing out at the helpless does not address the power imbalance between big tech companies and individual users of their products, people will be better equipped to recognize and refuse attempted manipulation by gender stereotype. At the same time, the current system of technology development cannot be trusted to fix itself.

As we reflect on how we *do* want to engage with voice assistants and what we might want them to be like, consideration of *li* in particular offers some helpful resources. First, if we are interested in quasi-social technologies that present at least some of the trappings of personhood, it will be important to keep in mind the difference between treating an AI as a person and attributing a gender to it. To the extent that gendering technologies—especially by using gender cues that are associated with widespread, robust, and pernicious stereotypes—interferes with our abilities to treat them with the respect and consideration we think is owed to persons, we have great reason to pay careful attention to existing social norms and the minimal cues that may be leveraged to activate them. We should also recognize that it is not enough to try *not* to activate gender schemas; we should be mindful about incorporating cues and prompts, as well as practices, that actively help us to *resist* them, whether that means activating gender cues in ways that subvert stereotypes (for instance, by associating femininity with competence and masculinity with warmth), identifying vocal ranges and names unlikely to activate gender schemas to begin with, or perhaps even emphasizing their nonhuman nature by encouraging users to think of them as analogous to, for example, animals or other nonhuman but personish entities that help create new imaginative possibilities, as Kate Darling (2021) argues.

There are many possibilities here—too many to assess in the present project. And even if we work out an appropriate *li* to pursue with virtual assistants given our current socially oppressive practices, this will not resolve the problem of what we might think of as phoniness, a propensity to follow the written rules without the appropriate emotional dispositions to make them genuine expressions of moral

concern. But I hope to have made the case that by reorienting ourselves to focus on the role of *li* in promoting humane social structures, we acquire helpful tools for thinking about both design and use considerations in the developing field of home AI, given the prevalence and power of gender schemas and the temptations to use and abuse them that are already being made apparent.

Acknowledgments

References

Addison, Arifah, Christoph Bartneck, and Kumar Yogeeswaran. 2019. "Robots Can Be More Than Black and White: Examining Racial Bias towards Robots." In *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 493–98. New York: Association for Computing Machinery. https://doi.org/10.1145/3306618.3314272.

Ames, Roger T. 2011. *Confucian Role Ethics: A Vocabulary*. Honolulu: University of Hawaiʻi Press.

Bartneck, Christoph, Kumar Yogeeswaran, Qi Min Ser, Graeme Woodward, Robert Sparrow, Siheng Wang, and Friederike Eyssel. 2018. "Robots and Racism." In *HRI '18: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 196–204. New York: Association for Computing Machinery. https://doi.org/10.1145/3171221.3171260.

Bogost, Ian. 2018. "Sorry, Alexa Is Not a Feminist." *Atlantic*, January 24, 2018. https://www.theatlantic.com/technology/archive/2018/01/sorry-alexa-is-not-a-feminist/551291/.

Borau, Sylvie, Tobias Otterbring, Sandra Laporte, and Samuel Fosso Wamba. 2021. "The Most Human Bot: Female Gendering Increases Humanness Perceptions of Bots and Acceptance of AI." *Psychology & Marketing* 38, no. 7 (July): 1052–68. https://doi.org/10.1002/mar.21480.

Cawston, Amanda. 2019. "The Feminist Case against Pornography: A Review and Re-evaluation." *Inquiry* 62 (6): 624–58. https://doi.org/10.1080/0020174X.2018.1487882.

Chalmers, M., I. MacColl, and M. Bell. 2003. "Seamful Design: Showing the Seams in Wearable Computing." In *IEE Eurowearable '03*, 11–16. London: Institution of Electrical Engineers. https://doi.org/10.1049/ic:20030140.

Confucius. 2003. *Confucius Analects: With Selection from Traditional Commentaries.* Translated by Edward Slingerland. Indianapolis: Hackett.

Crain, Matthew. 2018. "The Limits of Transparency: Data Brokers and Commodification." *New Media & Society* 20, no. 1 (January): 88–104. https://doi.org/10.1177/1461444816657096.

Darling, Kate. 2021. *The New Breed: What Our History with Animals Reveals about Our Future with Robots.* New York: Henry Holt.

Dreyfuss, Emily. 2018. "The Terrible Joy of Yelling at an Alexa." *Wired*, December 27, 2018. https://www.wired.com/story/amazon-echo-alexa-yelling/.

Eyssel, Friederike, and Frank Hegel. 2012. "(S)he's Got the Look: Gender Stereotyping of Robots." *Journal of Applied Social Psychology* 42, no. 9 (September): 2213–30. https://doi.org/10.1111/j.1559-1816.2012.00937.x.

Ferguson, Christopher J., and Richard D. Hartley. 2020. "Pornography and Sexual Aggression: Can Meta-Analysis Find a Link?" *Trauma, Violence, & Abuse* 23, no. 1 (January): 278–87. https://doi.org/10.1177/1524838020942754.

Fessler, Leah. 2017. "Siri, Define Patriarchy: We Tested Bots like Siri and Alexa to See Who Would Stand Up to Sexual Harassment." *Quartz*, February 22, 2017. https://qz.com/911681/we-testedapples-siri-amazon-echos-alexamicrosofts-cortana-and-googlesgoogle-home-to-see-which-personalassistant-bots-stand-up-for-themselvesin-the-face-of-sexual-harassment/.

Gustavsson, Eva. 2005. "Virtual Servants: Stereotyping Female Front-Office Employees on the Internet." *Gender, Work & Organization* 12, no. 5 (September), 400–419. https://doi.org/10.1111/j.1468-0432.2005.00281.x.

Kudina, Olya, and Mark Coeckelbergh. 2021. "'Alexa, define empowerment': Voice Assistants at Home, Appropriation and Technoperformances." *Journal of Information, Communication and Ethics in Society* 19 (2): 299–312. https://doi.org/10.1108/JICES-06-2020-0072.

Kupperman, Joel J. 2000. "Feminism as Radical Confucianism: Self and Tradition." In *The Sage and the Second Sex: Confucianism, Ethics, and Gender*, edited by Chenyang Li, 43–56. Chicago: Open Court.

LaBossiere, Michael. 2017. "Testing the Moral Status of Artificial Beings; Or 'I'm Going to Ask You Some Questions. . . .'" In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, edited by Patrick Lin, Ryan Jenkins, and Keith Abney, 293–306. Oxford: Oxford University Press.

Lai, Karyn L. (2008). *An Introduction to Chinese Philosophy.* Cambridge: Cambridge University Press.

Manne, Kate. 2020. *Entitled: How Male Privilege Hurts Women*. New York: Crown.

Markey, P. M., C. N. Markey, and J. E. French. 2015. "Violent Video Games and Real-World Violence: Rhetoric versus Data." *Psychology of Popular Media Culture* 4, no. 4 (October): 277–95. https://doi.org/10.1037/ppm0000030.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations.* Oxford: Oxford University Press.

Nass, Clifford, Youngme Moon, and Nancy Green. 1997. "Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers with Voices." *Journal of Applied Social Psychology* 27, no. 10 (May): 864–76. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x.

Ng, Alfred. 2019. "Amazon Alexa Keeps Your Data with No Expiration Date, and Shares It Too." *CNET*, July 2, 2019. https://www.cnet.com/home/smart-home/amazon-alexa-keeps-your-data-with-no-expiration-date-and-shares-it-too/.

Olberding, Amy. 2015. "From Corpses to Courtesy: Xunzi's Defense of Etiquette." *Journal of Value Inquiry* 49, no. 1–2 (March): 145–59. https://doi.org/10.1007/s10790-014-9466-5.

———. 2016. "Etiquette: A Confucian Contribution to Moral Philosophy." *Ethics* 126, no. 2 (January): 422–46. https://doi.org/10.1086/683538.

———. 2019. *The Wrong of Rudeness: Learning Modern Civility from Ancient Chinese Philosophy.* New York: Oxford University Press.

Otterbacher, Jahna, and Michael Talias. 2017. "S/he's Too Warm/Agentic! The Influence of Gender on Uncanny Reactions to Robots," In *HRI '17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 214–23). New York: Association for Computing Machinery. https://doi.org/10.1145/2909824.3020220.

Panzarino, Matthew. 2021. "Apple Adds Two Brand New Siri Voices and Will No Longer Default to a Female or Male Voice in iOS. *Tech Crunch*, March 31, 2021. https://techcrunch.com/2021/03/31/apple-adds-two-siri-voices/.

Puett, Michael. 2015. "Ritual and Ritual Obligations: Perspectives on Normativity from Classical China." *Journal of Value Inquiry* 49, no. 4 (December): 543–50. https://doi.org/10.1007/s10790-015-9524-7.

Puntoni, Stefano, Rebecca Walker Reczek, Markus Giesler, and Simona Botti. 2021. "Consumers and Artificial Intelligence: An Experiential Perspective." *Journal of Marketing* 85, no. 1 (January): 131–51. https://doi.org/10.1177/0022242920953847.

Rosenlee, Li-Hsiang Lisa. 2010. "A Feminist Appropriation of Confucianism." In *Confucianism in Context: Classic Philosophy and Contemporary Issues, East Asia and Beyond,* edited by Wonsuk Chang and Leah Kalmanson, 175–90. Albany: State University of New York Press.

———. 2016. Feminism and Multiculturalism Revisited: A Hybridized Confucian Care. In *The Bloomsbury Research Handbook of Chinese Philosophy and Gender*, edited by Ann A. Pang-White, 157–72. New York: Bloomsbury.

———. Forthcoming. *A Feminist Re-imagination of Confucianism: A Practical Ethic for Life*. New York: Columbia University Press.

Sap, Maarten, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. "The Risk of Racial Bias in Hate Speech Detection." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–78. Florence: ACL. https://doi.org/10.18653/v1/P19-1163.

Slingerland, Edward. 2014. *Trying Not to Try: Ancient China, Modern Science, and the Power of Spontaneity.* New York: Crown.

Steele, Chandra. 2018. "The Real Reason Voice Assistants Are Female (and Why It Matters)." *PC Magazine*, January 4, 2018. https://www.pcmag.com/opinions/the-real-reason-voice-assistants-are-female-and-why-it-matters.

Stroessner, Steven J., and Jonathan Benitez. 2019. "The Social Perception of Humanoid and Non-humanoid Robots: Effects of Gendered and Machinelike Features." *International Journal of Social Robotics* 11, no. 2 (April): 305–15. https://doi.org/10.1007/s12369-018-0502-7.

Táíwò, Olúfẹ́mi O. 2020. "Civility as Self-Determination." *Philosophy East and West* 70, no. 4 (October): 1073–83. https://doi.org/10.1353/pew.2020.0076.

US Equal Employment Opportunity Commission. n.d. *Diversity in High Technology Special Report.* https://www.eeoc.gov/special-report/diversity-high-tech.

West, Mark, Rebecca Kraut, and Han Ei Chew. 2019. *I'd Blush If I Could: Closing Gender Divides in Digital Skills through Education*. Document no. GEN/2019/EQUALS/1 REV 5. UNESCO and EQUALS. https://unesdoc.unesco.org/ark:/48223/pf0000367416.locale=en.

Withers, Rachel. 2018. "I Don't Date Men Who Yell at Alexa." *Slate: Future Tense*, April 30, 2018. https://slate.com/technology/2018/04/i-judge-men-based-on-how-they-talk-to-the-amazon-echos-alexa.html.

Wittkower, D. E. 2020. "What Is It Like to Be a Bot?" In *The Oxford Handbook of Philosophy of Technology,* edited by Shannon Vallor, 358–373. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190851187.013.23.

Wong, David. 2018. "Chinese Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021 edition. Published January 10, 2008; substantive revision September 14, 2018. https://plato.stanford.edu/archives/sum2021/entries/ethics-chinese/.

Wong, Pak-Hang. 2020. "Why Confucianism Matters for the Ethics of Technology." In *The Oxford Handbook of Philosophy of Technology,* edited by Shannon Vallor, 609–28. Oxford: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780190851187.013.36.

Xunzi. 2014. *Xunzi: The Complete Text.* Edited and translated by Eric L. Hutton. Princeton, NJ: Princeton University Press.

Zhu, Qin, Tom Williams, Blake Jackson, and Ruchen Wen. 2020. "Blame-Laden Moral Rebukes and the Morally Competent Robot: A Confucian Ethical Perspective." *Science and Engineering Ethics* 26, no. 5 (October): 2511–26. https://doi.org /10.1007/s11948-020-00246-w.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* London: Profile Books.

ALEXIS ELDER is an associate professor of philosophy at the University of Minnesota Duluth. Her research focuses on interpersonal relationships and new technologies, drawing on historical traditions to better understand emerging ethical issues. She is the author of *Friendship, Robots, and Social Media: False Friends and Second Selves* (Routledge 2018), as well as articles on topics ranging from companionate robots for geriatric care to chatbots of the dead.