Robotics has a race problem

Forthcoming, *Science, Technology, and Human Values* (accepted 7 June, 2019)

Professor Robert Sparrow
Department of Philosophy
School of Philosophical, Historical and International Studies
Monash University
ARC Centre of Excellence for Electromaterials Science
Monash University
Victoria 3800
Australia.

Abstract

If people are inclined to attribute race to humanoid robots, as recent research suggests, then designers of social robots confront a difficult choice. Most existing social robots have white surfaces and are therefore, I suggest, likely to be perceived as White, exposing their designers to accusations of racism. However, manufacturing robots that would be perceived as Black, Brown, or Asian, risks representing people of these races as slaves, especially given the historical associations between robots and slaves at the very origins of the project of robotics. The only way engineers might avoid this ethical and political dilemma is to design and manufacture robots to which people will struggle to attribute race. Doing so, however, would require rethinking the relationship between robots and "the social" that sits at the heart of the project of social robotics. Discussion of the race politics of robots is also worthwhile because of the potential it holds to generate insights about the politics of artifacts, the relationship between culture and technology, and the responsibilities of engineers.

Keywords: Robotics; Ethics; Robots; Race; Racism; Diversity; HRI; Media

Robotics has a race problem

Introduction

Recent progress in robotics technology has once again led people to predict that robots will soon enter our homes in significant numbers and our lives on a daily basis. As people have begun to take this possibility seriously, they have started to wonder about what it would mean to live alongside of, and in relationship with, robots. How would this affect our relationships with each other? What choices will we need to confront as we design robots to interact with human beings? And what ethical issues will arise as we confront them (Dumouchel and Damiano 2017; Lin et al. 2012)?

This paper makes a contribution to this larger project by highlighting a particular ethical and political dilemma that may arise when engineers build humanoid social robots of the sort that they hold will share our homes and workplaces in the future.[1] The origins of the dilemma lie in the tendency people have to attribute various traits to robots based on their experience with humans and animals. If one of the things that people attribute to some robots is race, as recent research suggests, then designers of social robots confront a difficult choice. Most existing social robots have white surfaces and are therefore, I suggest, likely to be perceived as White, exposing their designers to accusations of—arguably mostly unintentional—racism.[2] However, manufacturing robots that would be perceived as Black, Brown, or Asian,[3] risks representing people of these races as slaves, especially given the historical associations between robots and slaves at the very origins of the project of robotics. The only way that I can see engineers might avoid this ethical and political dilemma is to design and manufacture robots to which people will struggle to attribute race. Doing so, however, would require rethinking the relationship between "the social" and robots that sits at the heart of the project of social robotics. Discussion of the race politics of robots is also worthwhile because of the potential it holds to generate insights about the politics of artifacts, the relationship between culture and technology, and the responsibility of engineers.

The structure of my discussion is as follows. In Section I, I note the ways in which people are willing to attribute various traits to robots; I draw attention to recent evidence that suggests

that one of the traits they attribute is race; and I defend the idea that we might plausibly think of robots as "having race." In Section II, I argue that, if robots have race, the vast majority of real social robots are White, and defend this claim against a number of objections. Section III explains why the race of robots should matter to engineers, both morally and pragmatically. In Section IV, I use the work of Louis Chude-Sokei (2016) to explore the reasons why the obvious response to the current racial homogeneity of social robots—that is, creating a more racially diverse range of social robots—is itself likely to be politically problematic. I also offer some brief thoughts about a possible way forward and the challenge this poses for the project of social robotics. At the conclusion of my treatment I draw out some of the implications of my discussion for Science and Technology Studies (STS) scholarship more generally.

I. How robots have race

When people interact with robots they have a strong tendency to anthropomorphize them (Ishiguro and Nishio 2018; Phillips et al. 2018; Reeves and Nass 1998). They adopt an "intentional stance" (Dennett 1987) and attribute desires and beliefs to robots (Krach et al. 2008; Lee et al. 2005; Powers and Kiesler 2006). People's expectations of, and behavior towards, robots are also shaped by their appearances (Bartneck et al. 2009; Haring et al. 2016; Hegel et al. 2008), which suggests that we place robots in the same social and ontological schemas that we use to classify people and animals. Thus, for instance, people have different expectations of humanoid robots than they do of robots shaped like machines (Hegel et al. 2008; Kwak 2014) or animals (Lee et al. 2011), and can quickly identify a robot as being a robot dog or a robot cat or an android.[4] Similarly, people are remarkably quick to attribute gender to robots, and their relationships with robots are shaped by whether they are interacting with a "male" or "female" robot (Bernotat et al. 2017; Eyssel and Hegel 2012; Otterbacher and Talias 2017; Robertson 2018; Siegel et al. 2009).

Given these common responses to robots it is hardly surprising that recent research suggests that, as well as attributing gender and species to robots, people also attribute race to robots (Bartneck et al. 2018; Eyssel and Loughnan 2013). That is to say, people are willing to identify robots as being White or Black or Asian, and their judgments about these matters

demonstrate a high degree of inter-subjective reliability. More controversially, there is some evidence that people's interactions with robots are shaped by the race of the robot whether they *consciously* identify it as having a race or not (Bartneck et al. 2018).[5]

It might be objected that robots cannot have race because race is something that only human beings have. There is some truth in this claim. There are aspects of race that robots are unlikely ever to possess. In particular, unless they become sentient, robots will not have the lived experience of racism, nor an individual life history that implicates them in a racial politics.[6]

Nevertheless, there are a number of reasons to believe robots *do* have race.

To begin with, note that language is always (more obviously) metaphoric when people are talking about robots. Do robots have arms, legs, and heads? Do they have faces? They do not, after all, have these things like we do. Their heads are usually plastic and may be of any shape; their legs are metal instead of flesh and blood. Yet people have no trouble at all, if asked to do so, in pointing to a robot's head or legs. Even engineers and roboticists who are all-too-aware of the differences between robots and people are quite comfortable in talking about the legs of the robot. Indeed, the development and cultural currency of robots has arguably expanded our sense of what counts as a leg so that a robot "leg" really is, now, a leg. At the very least, robots have race in this sense: People can identify what race a robot is, and their responses to the robot are shaped by what it is that they identify (Bartneck et al. 2018; Eyssel and Loughnan 2013). If this race is only metaphoric ("race"), it is no more or less so than the rest of the anthropomorphic concepts that we routinely apply to robots.

Moreover, according to an influential account of the nature of race, the relationships that people have with robots may be sufficient to establish that some robots genuinely do have race. "Social constructionist" scholarship about race insists that there is ultimately nothing more to race than the way people respond to each other (Appiah 1996; Zack 2002, 87-88). Race is neither ancestry, skin color, ethnicity, nor culture (Templeton 2013). The social categories of race do not correspond with the patterns of statistical associations between people descended from ancestors from particular geographic regions that researchers can detect at the level of genetics (Knowles 2010, 31; Smedley and Smedley 2005, 22; Templeton 2013). Indeed, there is nothing that all members of a race share with each other

except the fact that they are all treated as being members of that race. Instead, race is a matter of identification with a particular group of persons and of being identified as a member of that group (Denton and Deane 2010). Recognizing a person's race is to apply a social framing and to mentally—and sometimes socially—place them in a relationship with other people (Root 2000). For instance, Blacks are associated with each other and contrasted with Whites; Whites are thought to share something in common that is not shared by Blacks or Asians, *et cetera*. These identifications and beliefs in turn are sufficient to generate the other social regularities that further constitute race (Knowles 2010, 30-33; Root 2000; Smedley and Smedley 2005). Thus, as a result of being identified as Black, those individuals who are so identified may come to have distinctive experiences which, in turn, cause them to have more in common with each other and to respond in particular ways that become associated with that identification.

On this account of the nature of race, then, race is always a function of representation: It exists at the level of meaning or signification rather than biology.[7] If people did not recognize race in each other, race would cease to exist.[8] Conversely, race exists wherever people recognize it. This means that if people attribute race to robots and respond to them as if they have race then robots really do have race. Knowing the race of the robot—how people locate it within the local schema of racial categories—will help predict how people respond to it in various contexts in which race is salient. As I argue further below, this means that in some contexts the race of the robot might turn out to be very important.

II. Robotics has a race problem

The fact that (some) robots might have race is not in and of itself a problem, although as I will argue further below, the historical legacy of racism means that any representation of race by a robot is *potentially* problematic. What is clearly a problem, however, is the fact that a large majority of humanoid robots are white in color and therefore arguably White, as can be demonstrated by reference to the "anthropomorphic robots database" (Phillips et al. 2018).[9] This database contains images of robots from research laboratories around the world today and ranks them according to their "human-likeness" score. Eliminating the most obviously non-humanoid robots, by excluding robots with a human likeness score of 24 (out of 100) or below, generates, at the time of writing, a sample of 125 humanoid

robots. Of these, 61 have surfaces that are entirely white, 21 are almost entirely white, with a small splash of another color, eight are mostly white but a significant portion of their surface is another color, sixteen are predominantly silver, fourteen are predominantly blue, red, green, yellow, or another bright color, two are predominantly grey, and only three have predominantly black or brown surfaces. Of this sample, then, between 66% and 72% seem likely to be racially coded as White.[10] Moreover, because most of these robots are research robots—rather than robots that are commercially available—these figures significantly underestimate the percentage of those robots that are actually manufactured that are white. Both of the most popular commercially available humanoid social robots are white: Pepper is white and Nao is white.[11] If people encounter a humanoid social robot in their daily lives, then, it will almost always be white.

There are a number of exceptions to the rule that robots are racially coded as White, which only serve to prove it.

First, there are a small number of widely publicized robots, the appearance of which is modeled on that of particular individuals who are themselves non-White. For instance, Professor Hiroshi Ishiguro is famous for having built realistic looking animatronic "robots" in the image of himself and also his wife (Guizzo 2010). These robots have the same Asian appearance as Prof Ishiguro and his spouse. Similarly, Martine Rothblatt and Hanson Robotics collaborated to design and manufacture an animatronic talking head, modeled on Martine Rothblatt's wife, Bina Rothblatt, which they call Bina48 (Harmon 2010). Because Bina Rothblatt is African-American, Bina 48 appears Black. In these cases the representation of race by the robot plays a crucial role in establishing that it represents the desired individual.

Second, there are robots that have been developed for purposes of mobilizing national pride in the developer's nation's scientific and technological prowess and/or specifically to cater to local cultural mores. For instance, researchers at the Interactive Robots and Media Laboratory at United Arab Emirates University built what they advertised as the "first humanoid robot which actually can perform Arabic dialogue," (Gonn 2009) which they named Ibn Sina, after an 11th Century Muslim philosopher (Riek et al. 2010). A news story about this robot includes video footage in which the robot appears with olive skin and a

long grey beard and is wearing a turban (Gonn 2009). Here the robot has been racialized to be non-White in order to convey a national or regional identity. Its race allows it to serve as a symbol of national scientific prowess and technical ingenuity, as well as reassure a local audience that robots need not necessarily be a foreign and "Western" technology.

Finally, there is at least one instance of a robot being designed and promoted as an explicit racist caricature. In 1930, in the United States, Westinghouse Electric and Manufacturing Co manufactured and promoted what they called the "Rastus robot" (Johnson 2016). This machine had black rubberized skin and Negroid features and was displayed and publicized in a number of contexts that made it clear that the robot was a grotesque figure intended to play the role of a Black slave or servant. Indeed, Simone Browne has argued that the ideological function of the Rastus robot was in fact to convince White manufacturing workers that they would maintain their racial privilege even in a world in which robots started to threaten their jobs (Browne 2017). Rastus could serve in this role because "he" was very clearly a Black robot.

These non-White robots suggest that the vast majority of contemporary humanoid robots are White in two ways. First, they make it clear that it is possible for robots to have race and that the race of a robot can be altered by changing a few surface features of the robot, such as its "skin" and "hair." If these robots succeed in having race, then so does the mainstream of white robots. Second, the fact that these examples stand out, and that it is necessary to alter the appearance of robots in order to establish them as being non-White, shows that the "default" race of humanoid robots today is indeed White.[12]

It might be objected that the white plastic surfaces that are typical of contemporary robots fail to establish that they are White. The predominant social imaginary of the future, established in science fiction (especially film) but also in advertising, is characterized by gleaming metallic and white plastic surfaces (Bernardi 1998, 71-76; Banham 2010, 166; Conekin 2010, 146-149). We fantasize that the future will be new, clean, and sparkling, and in images of the future the absence of dirt or wear is highlighted by the white and silver surfaces of the objects therein. It might therefore be argued that the reason why designers provide their robots with white (and silver) surfaces, then, is not in order to establish that

they are White but rather in order to invoke this imaginary and establish that they represent the future.

However, the intentions of designers are not determinative when it comes to the question of whether robots have race: people may attribute race to robots that were designed by people who never thought of them as having race. If they do—and, as we have seen, there is evidence that they do—this will mean that robots do have race regardless of the intentions of their designers.

Moreover, to object that robots cannot be White because they do not have realistic skin color, is to misunderstand the nature of Whiteness, which is only incidentally about the color of "White" people's skin (Mills 1997, 127; Painter 2010, ix).[13] White skin is not, after all, white; at most it is closer to white than the skin of individuals identified with other races (Painter 2010, 394). Indeed, there is a sense in which, unlike individuals of other races, White people are not thought of as having color. Instead, in the classical European understanding of race, White bodies are the defaults, the "original" or paradigm human bodies, variations on which produce bodies that do have race (Dyer 1997, 1-4). Other races are "colored," then, but to be white is not to be defined by the color of one's skin but rather by reference to the imaginary possession of a range of virtues and properties that are associated with Whiteness, such as purity, strength, and self-reliance (Roediger 2005, 63-65, 70-78; Dyer 1997). Moreover, as critics—as well as the advocates and authors of the alternative science fiction aesthetic, Afrofuturism—have highlighted, these ideas about Whiteness have played a crucial role in shaping the traditional science fiction aesthetic, which itself then reflects and reinforces them (Bernardi 1998, 75-80, 102-104; Davis 1983; Dinerstein 2006, 578-581; Nama 2009, 156, 159, 164; Womack 2013, 5-7). The social contamination—the "dirt," if you will—that the aesthetic denies is in part the presence of people of color (Bernardi 1998, 75-76, 86-89; Dyer 1997, 65-66, 75-76).[14] Thus, in fact, the glossy pure white surfaces of robots are more effective in signaling the Whiteness of robots than providing them with a more naturalistic skin tone. We might even say that robots are *more* White than White people.

Because Whiteness does not require whiteness, it is possible that even robots with silver and gold surfaces are White. This is an empirical matter and it will depend upon what race, if any, people are inclined to assign to such robots.

One possibility is that silver or gold anthropomorphic machines may acquire race by virtue of their morphology, accents, behavior, or social roles. Robots with stereotypically racialized facial features (or "hair") may be assigned race on the basis of these features even where the color of their surfaces makes the assignation incongruous. In some contexts, a manner of speaking and/or an accent can communicate and imply race (Anderson 2007; Lippi-Green 2012, 182; Rakić et al. 2011). Finally, in societies where particular occupations or social roles tend to be performed by persons of particular races, people may respond to the role of the robot as a racial cue and attribute race accordingly (Freeman et al. 2011). Robots that do not initially appear white may nevertheless come to be understood as White as a result of any of these processes.

However, it is also possible that silver or gold robots are racially coded White by virtue of the chain of imaginary associations mentioned above, or simply because the alternatives of assigning them to another race or suspending our ingrained habits of assigning race to people are more incongruous and cognitively demanding than acknowledging them to be White.

Nevertheless, it is possible that people will *not* identify robots with silver or gold surfaces as White or respond to them as if they are White. Moreover, it seems almost certain that some robots will fail to have race. People do not typically attribute race to animals, for instance, or to their toasters or washing machines. Social robots that do not look like human beings or invoke the same sets of social categories that we use in relating to human beings in our responses to them are unlikely to have race attributed to them.

It is important to understand, therefore, that in claiming that some robots might have race I am making an empirical claim about how people respond to humanoid social robots (as well as a socio-theoretic claim about how we should interpret their responses) rather than a claim that robots *necessarily* have race. However, it is equally important to understand that I am not allowing that those robots that do have race do not *really* have race or that the race of robots is only "in people's heads." I want to insist that when a robot has race, its

race is a real feature with causal powers. In particular, any social robot that learns from experience is likely to internalize the racial concepts that people use in their interactions with the robot. Just as, in societies where race matters, children "learn race" as they grow up (Feagin and Van Ausdale 2001), so too will learning robots come to possess racial identities through their history of social interactions if people attribute race to them. If we imagine two otherwise identical robots, then, one white and one black, in a society divided into Blacks and Whites, and people understand the robots to have race, then we should anticipate that they might eventually behave quite differently. In this way, people perceiving robots to have race will generate real differences in the behavior of the robots as social agents. Nor is it necessary in order for a robot to have race that every person who encounters it recognizes it as doing so. Rather, robots have race when the majority of people who encounter them respond to them in accordance with a racial schema.[15] Like people, then, robots may have race even when particular people do not recognize it.

III. Why the race of robots matters

One reason I have described the overwhelmingly White racial coding of social robots as a race *problem* relates to a traditional concern about a lack of diversity in the media. When members of a minority community do not see themselves represented in the media, or when those representations that do exist are negative, both racial equality and social solidarity are undermined (Whitehouse 2009). Insofar as a lack of diversity of representations either reflects a morally pernicious blindness to the presence of diversity in the community or the deliberate choice not to include representations of members of racial minorities, it is also hard not to think that the failure to acknowledge racial diversity in the media is itself reflective—as well as an expression—of racism.

Thus, an important reason why engineers should become more conscious of the race of the robots they build is to avoid accusations of racism of this sort. Note that while producing a more racially diverse range of robots might be a necessary step to avoid accusations of racism, it will not be sufficient: designers will also need to be sensitive to the ways in which placing robots of particular colors in particular roles is likely to be received. If educational robots are always White, but the shoeshine robots are always Black and the robot butlers

always Brown, this will reinforce pernicious racial narratives. Engineers have, I strongly believe, a moral obligation to avoid and resist such stereotypes.

However, there is another, pragmatic, reason why the designers of social robots should be thinking more about race. In societies that are deeply structured by race, as is the United States today, race affects how people communicate (Bell and Johnson 1997; Rickford and McNair-Knox 1994; Rickford and Rickford 2000), how much they trust each other (Smith 2010; Stanley et al. 2011), and how they treat each other (Dasgupta 2004; Fiske 1998; Rooth 2010). If, as I have argued, robots have race, the same thing will be true for interactions between people and robots. If the designers of social robots wish to facilitate certain sorts of responses to, or engagements with, their robots, then, they will need to take these dynamics into account. For instance, because both the amount of time people spend communicating and their satisfaction with that communication is partially determined by racial dynamics (Cooper et al. 2003; Johnson et al. 2004), wherever it is important that people trust a robot—which includes, especially, healthcare and educational settings—designers will need to consider the possibility that they will need robots of different races in different contexts. The more the sorts of social interactions for which the robot is designed are structured by race the more the race of robots will matter. If we imagine robots taking on roles in law enforcement in the United States, for example, then the race of the robots is likely to matter a lot. African-Americans may be less willing to trust a White robot and sending a White robot to police a Black neighborhood may be seen as provocative. Unless the designers of social robots pay attention to the race politics of their robots, there is a risk that their robots will fail badly in many applications.

IV. Robotics has an ethical problem

I have argued that humanoid social robots are likely to represent people of different races, typically White people. My purpose in doing so thus far has been to emphasize the way in which social robots may have *race*. However, the fact that such robots also represent *people* also turns out to be important in this context. By and large humanoid robots will be designed to be valuable to people by serving them in various ways. Robots will respond to human requests and will do what people tell them to do. That is to say, robots will essentially be slaves (Bryson 2010) and will, therefore, represent people as slaves. Indeed,

the idea that robots are mechanical versions of human slaves is present from the very beginning of robotics: the word "robot" itself is derived from the Czech word "robota," meaning worker or slave.

There is, I think, a real question about the ethics of representing people as slaves.[16] One possibility is that this will affect our behavior towards other people so that we are more inclined to treat them as slaves.[17] It is also plausible that representing people as slaves expresses a morally problematic disrespect for them regardless of whether or not it shapes our behavior towards each other. Unfortunately, I do not have the space to make either of these arguments at length here.[18]

Yet representing Black or Brown people as slaves is clearly especially morally problematic given the history of slavery in the United States and elsewhere. This means that efforts to introduce racial diversity into the range of social robots are going to confront a difficult problem. Social robots will be unrepresentative of the broader community and risk further exacerbating existing racism in the ways discussed above unless designers take steps to increase the diversity in the surface colors and (perhaps) facial structures of the robots. However, if they do manufacture robots that look Black or Brown, then these robots will represent Black slaves or Brown servants, with all the troubling historical resonances that go along with that.

This dilemma is exacerbated by a further subtlety of the cultural politics of robots. In my account of the semiotics of the traditional science fiction aesthetic I emphasized how the color scheme associated with the future works to deny the presence, in the future, of people of color. However, there is another way in which racism has shaped the European vision of the future, especially in relation to robots. As Louis Chude-Sokei has emphasized, the fact that, from the very beginning of their history, robots have been thought of as mechanical slaves means that the history of slavery as well as assumptions about slavery have played a role in shaping our understanding of robots (Chude-Sokei 2016, 85-128). Robots were supposed to have the same sets of "virtues" (hard-working, obedient, servile) as slaves and perform the same sets of tasks. Representations of robots in film and literature have almost always functioned to express or work through social anxieties about either the institution of slavery or the end of the White privilege that it sustained (Morrell

2015). Thus, while the white surfaces of contemporary social robots suggest that people will be inclined to perceive them as White, the idea that they are in fact Black or Brown is, as a result of this history, all-too-readily available.[19]

Social robotics therefore faces a difficult ethical as well as a political dilemma when it comes to the question of the race of humanoid robots. The status quo, in which these robots are almost universally racially coded as White, is indefensible. Yet the obvious response to the current racial homogeneity of social robots—creating a more racially diverse range of social robots—is itself likely to be morally problematic insofar as it will involve reproducing and reinforcing traditional racist ideas about race and servitude.

The only way around this dilemma that I can see would be to try to design robots to which people do not attribute race. One option would be to make social robots blue or green or some other combination of colors that makes it difficult for people assign race to them. Further research on why, and in what circumstances, people are inclined to attribute race to robots will be necessary to be confident that this option is available (Bartneck et al. 2018; Sparrow 2019). It is possible, for instance, that the cultural presuppositions identified by Chude-Sokei mean that any non-white humanoid robot in a service role, no matter what its color, is effectively racially coded as Black or Brown.

Another option would be to make social robots that were not remotely humanoid, again in the hope that people will not attribute race to them. There is a long-standing and influential tradition of thought in robotics that holds that the best way to make robots "social" is to build humanoid robots (Breazeal 2003, 119-124; Duffy 2003, 177; Hegel et al. 2208; Kwak 2014; Phillips et al. 2018). If robots are to assist us in our daily lives, they will need to be able to navigate and to fit into homes and workplaces, which are designed for the human form. The human tendency to anthropomorphize robots means that humanoid robots are able to make use of non-verbal cues to communicate more effectively than robots that look more obviously like machines and to elicit human emotions to help shape social interactions. However, my discussion suggests that in order to avoid the complex politics of the social category of race impacting negatively on the performance of robots in social interactions, it may be necessary to build machines that foreground the fact that they are *not* human.[20]

More generally, the dialectic explored above highlights the fact that social meanings with which social robots and their designers must engage will not always work to the advantage of either. In particular, the existence of widespread sexism and racism in many of the societies in which robots will be operating is likely to have profound implications for both the effectiveness of particular sorts of robots in particular roles and the ethics of their design. Engineers need to be thinking more about these matters now to reduce the risk that the machines they build will embed the social injustices of today in the years to come.

V. Robots, race, and STS

My main purpose in this essay has been to draw attention to the ethical and political dilemma posed by the possibility that robots have race. However, I believe that discussions about the race politics of robots also have significant potential to shed light on, and generate new insights in, long-standing controversies in STS scholarship.

For instance, how the design and function of humanoid robots seems to be in a relationship with the history of slavery is a particularly striking example of the sense in which artifacts "have politics" (Winner 1980). The fact that humanoid robots refer to, and represent, human beings means that their design as machines intended to serve refers to the idea of human slaves. It is the real-world history of slavery and race relations that gives them the particular politics they have today but it's difficult to see how there could be any world in which humanoid robots would not raise ethical and political questions about slavery. The complex interaction between design, function, and history in determining the politics of robots makes them a productive site to investigate questions about the politics of technology more generally.

Similarly, robots represent an under-utilized opportunity to further interrogate the role played by literature, film, and television in shaping technologies and their reception. While there has already been some research done on both the cultural politics of gender in robotics (Robertson 2018) and the intersection of ideas about race and technology in robots (Chude-Sokei 2016; Hampton 2015; Kakoudaki 2014), more work remains to be done to draw out the lessons of this research for the larger debate about how the arts shape technology (and vice versa). In particular, are there general lessons that might be learned

about when and how the cultural coding of a technology may generate ethical dilemmas for designers?

Finally, the ethical and political dilemma arising the race politics of robots raises questions about the responsibilities of scientists and engineers (Jasanoff 2016). As noted above, robots may effectively "have race" even when engineers have never thought about the matter. Indeed, people may attribute race to robots *against* the intentions of engineers. When the question of the responsibility of engineers arises, usually at issue is their responsibility for the uses of the technologies they design. However, in this case a question also arises about the responsibility of designers for *the meanings that are attributed to* the technologies that they design. Engineers face a dilemma similar to that faced by authors and artists regarding the reception of their work: to what extent are they responsible for the way other people interpret it? This question becomes more and more important for engineers in a political climate where the politics of speech and the ethics of representation are increasingly harmful and contested and new technologies are increasing oriented towards "the social."

Acknowledgements

The research for this paper was supported under the Australian Research Council's Centres of Excellence funding scheme (project CE140100012). The views expressed herein are those of the authors and are not necessarily those of the Australian Research Council. Some of the research for this paper was conducted while I was a Visiting Fellow in the Department of Philosophy at Carnegie Mellon University: I would like to thank David Danks for the opportunity to visit CMU. Thanks are also owed to Jess Berry, Simone Browne, Louis Chude-Sokei, Frederic Gilbert, Linda Hancock, Mara Harrell, Howie Manns, Andrew Milner, Sascha Morrell, Illah Nourbakhsh, Joshua Specht, and Aaron Steinfeld, for discussions and/or correspondence while I was drafting the essay. Mark Howard and Courtney Hempton assisted with preparing the manuscript for publication.

Author Biography

Rob Sparrow is a Professor in the Philosophy Program, a Chief Investigator in the Australian Research Council Centre of Excellence for Electromaterials Science, and an adjunct Professor in the Monash Bioethics Centre, at Monash University, where he works on ethical issues raised by new technologies. He is a co-chair of the IEEE Technical Committee on Robot Ethics and was one of the founding members of the International Committee for Robot Arms Control.

Bibliography

Anderson, Kate. 2007. "Constructing “otherness”: ideologies and differentiating speech style." *International Journal of Applied Linguistics* 17 (2):178-197.

Appiah, K Anthony. 1996. "Race, culture, identity: Misunderstood connections." In *Color conscious: The political morality of race*, edited by K Anthony Appiah and Amy Gutmann, 30-105. Princeton: Princeton University Press.

Banham, Reyner. 2010. "All that glitters is not stainless." In *The Design History Reader*, edited by Grace Lees-Maffei and Rebecca Houze, 165-169. Oxford: Berg Publishers.

Bartneck, Christoph, Takayuki Kanda, Omar Mubin, and Abdullah Al Mahmud. 2009. "Does the design of a robot influence its animacy and perceived intelligence?" *International Journal of Social Robotics* 1 (2):195-204.

Bartneck, Christoph, Kumar Yogeeswaran, Qi Min Ser, Graeme Woodward, Robert Sparrow, Siheng Wang, and Friederike Eyssel. 2018. "Robots and racism." *2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI ’18)*, New York: ACM.

Bell, Allan, and Gary Johnson. 1997. "Towards a sociolinguistics of style." *University of Pennsylvania Working Papers in Linguistics* 4 (1):2.

Bernardi, Daniel. 1998. *Star Trek and history: Race-ing toward a white future*. New Brunswick: Rutgers University Press.

Bernotat, J, Friederike Eyssel, and J Sachse. 2017. "Shape It – The influence of robot body shape on gender perception in robots." In *Social Robotics, ICSR 2017*, edited by A Kheddar, E Yoshida, S.S Ge, K Suzuki, J.-J Cabibihan, Friederike Eyssel and H He, 75-84. Cham: Springer.

Breazeal, Cynthia. 2003. "Emotion and sociable humanoid robots." *International Journal of Human-Computer Studies* 59 (1-2):119-155.

Browne, Simone. 2017. Thinking A.I. through Rastus Robot, the Westinghouse Mechanical Slave. In *AI Now Experts Workshop 2017*. New York.

Bryson, J.J. 2010. "Robots should be slaves." In *Close engagements with artificial companions: key social, psychological, ethical and design issues*, edited by Yorick Wilks, 63-74. John Benjamins Publishing.

Butler, Judith. 1990. *Gender Trouble : Feminism and the Subversion of Identity*. New York: Routledge.

Chude-Sokei, Louis. 2016. *The sound of culture: Diaspora and black technopoetics*. Middletown: Wesleyan University Press.

Conekin, Becky. 2010. "'Here is the modern world itself', the festival of Britain's epresentations of the future." In *The Design History Reader*, edited by Grace Lees-Maffei and Rebecca Houze, 143-151. Oxford: Berg Publishers.

Cooper, Lisa A, Debra L Roter, Rachel L Johnson, Daniel E Ford, Donald M Steinwachs, and Neil R Powe. 2003. "Patient-centered communication, ratings of care, and concordance of patient and physician race." *Annals of internal medicine* 139 (11):907-915.

Crawford, Kate. 2016. "Artificial intelligence's white guy problem." *The New York Times*, June 25.

Dasgupta, Nilanjana. 2004. "Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations." *Social Justice Research* 17 (2):143-169.

Davis, Thulani. 1983. The Future May Be Bleak, But It's not Black. *The Village Voice*, February 1, 17-20.

Dennett, Daniel Clement. 1987. *The intentional stance*. Cambridge: MIT press.

Denton, Nancy A, and Glenn D Deane. 2010. "Researching race and ethnicity: Methodological issues." In *The SAGE handbook of race and ethnic studies*, edited by John Solomos and Patricia Hill Collins, 67-89. London: Sage Publications.

Dinerstein, Joel. 2006. "Technology and its discontents: On the verge of the posthuman." *American Quarterly* 58 (3):569-595.

Duffy, Brian R. 2003. "Anthropomorphism and the social robot." *Robotics and autonomous systems* 42 (3-4):177-190.

Dumouchel, Paul, and Luisa Damiano. 2017. *Living with robots*. Translated by Malcolm DeBevoise. Cambridge: Harvard University Press.

Dyer, Richard. 1997. *White: Essays on race and culture* New York: Routledge.

Eyssel, Friederike, and Frank Hegel. 2012. "(S)he's got the look: Gender stereotyping of robots." *Journal of Applied Social Psychology* 42 (9):2213-2230.

Eyssel, Friederike, and Steve Loughnan. 2013. "It don't matter if you're black or white?" In *Social Robotics. ICSR 2013*, edited by G Herrmann, M.J Pearson, A Lenz, P Bremner, A Spiers and U Leonards, 422-431. Cham: Springer.

Feagin, Joe R, and Debra Van Ausdale. 2001. *The first R: How children learn race and racism*. Lanham: Rowman & Littlefield Publishers.

Fiske, Susan. 1998. "Stereotyping, prejudice, and discrimination." In *The Handbook of of Social Psychology*, edited by DT Gilbert, Susan Fiske and G Lindsay, 357-411. Boston: McGraw Hill.

Freeman, Jonathan B, Andrew M Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady. 2011. "Looking the part: Social status cues shape race perception." *PloS one* 6 (9):e25107.

Garland-Thomson, Rosemarie. 2017. *Extraordinary Bodies : Figuring Physical Disability in American Culture and Literature*, 20th Anniversary Edition. New York: Columbia University Press.

Gonn, Adam. 2009. "World's first Arabic speaking robot constructed in UAE." *The Jerusalem Post*, November 4.

Guizzo, Erico. 2010. "The man who made a copy of himself." *IEEE Spectrum* 47 (4):44-56.

Hampton, Gerome J. 2015. *Imagining Slaves and Robots in Literature, Film, and Popular Culture: Reinventing Yesterday's Slave With Tomorrow's Robot*. London: Lexington Books.

Haring, Kerstin S., David Silvera-Tawil, Tomotaka Takahashi, Katsumi Watanabe, and Mari Velonaki. 2016. "How people perceive different robot types: A direct comparison of an android, humanoid, and non-biomimetic robot." In *8th International Conference*

on Knowledge and Smart Technology (KST), edited by Mahāwitthayālai Būraphā, 265-270. New York: IEEE.

Harmon, Amy. 2010. "Making Friends with a robot named Bina48." *The New York Times*, July 4.

Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.

Hegel, Frank, Soren Krach, Tilo Kircher, Britta Wrede, and Gerhard Sagerer. 2008. "Understanding social robots: A user study on anthropomorphism." In *RO-MAN 2008: The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 574-579. New York: IEEE.

Howard, Ayanna, and Jason Borenstein. 2018. "The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity." *Science and Engineering Ethics* 24(5):1521-1536.

Ishiguro, Hiroshi, and Shuichi Nishio. 2018. "Building Artificial Humans to Understand Humans." In *Geminoid Studies: Science and Technologies for Humanlike Teleoperated Androids* edited by Hiroshi Ishiguro and Fabio Dalla Libera, 21-37. Singapore: Springer Singapore.

Jasanoff, Sheila. 2016. *The ethics of invention: technology and the human future*. New York, W.W. Norton & Company.

Johnson, Bob. 2016. "Energy slaves: Carbon technologies, climate change, and the stratified history of the fossil economy." *American Quarterly* 68 (4):955-979.

Johnson, Rachel L, Debra Roter, Neil R Powe, and Lisa A Cooper. 2004. "Patient race/ethnicity and quality of patient–physician communication during medical visits." *American journal of public health* 94 (12):2084-2090.

Kakoudaki, Despina. 2014. *Anatomy of a Robot: Literature, Cinema, and the Cultural Work of Artificial People*. New Brunswick: Rutgers University Press.

Knowles, Caroline. 2010. "Theorising race and ethnicity: contemporary paradigms and perspectives." In *The SAGE handbook of race and ethnic studies*, edited by John Solomos and Patricia Hill Collins, 23-42. London: Sage Publications.

Krach, Sören, Frank Hegel, Britta Wrede, Gerhard Sagerer, Ferdinand Binkofski, and Tilo Kircher. 2008. "Can machines think? Interaction and perspective taking with robots investigated via fMRI." *PloS one* 3 (7):e2597.

Kwak, Sonya S. 2014. "The Impact of the Robot Appearance Types on Social Interaction with a Robot and Service Evaluation of a Robot." *Archives of Design Research* 27(2): 81-93.

Lavender III, Isiah. 2014. *Black and brown planets: The politics of race in science fiction*. Jackson: University Press of Mississippi.

Lee, Sau-lai, Ivy Yee-man Lau, and Ying-yi Hong. 2011. "Effects of appearance and functions on likability and perceived occupational suitability of robots." *Journal of Cognitive Engineering and Decision Making* 5 (2):232-250.

Lee, Sau-lai, Ivy Yee-man Lau, Sara Kiesler, and Chi-Yue Chiu. 2005. "Human mental models of humanoid robots." *2005 IEEE International Conference on Robotics and Automation*, Barcelona: IEEE, 2767-2772.

Levy, David. 2009. "The ethical treatment of artificially conscious robots." *International Journal of Social Robotics* 1 (3):209-216.

Lin, Patrick, Keith Abney, and George A Bekey. 2012. *Robot ethics: the ethical and social implications of robotics*. Cambridge: MIT press.

Lippi-Green, Rosina. 2012. *English with an accent: Language, ideology, and discrimination in the United States*. New York: Routledge.

Mills, Charles, W. 1997. *The Racial Contract*. Ithaca: Cornell University Press.

Morrell, Sascha. 2015. "Zombies, robots, race, and modern labour." *Affirmations: of the modern* 2 (2):101-134.

Nama, Adilifu. 2009. "R is for race, not rocket: Black representation in American Science fiction cinema." *Quarterly Review of Film and Video* 26 (2):155-166.

O'Neill, Catherine. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. London: Allen Lane.

Otterbacher, Jahna, and Michael Talias. 2017. "S/he's too warm/agentic!: The influence of gender on uncanny reactions to robots. " *2017 ACM/IEEE International Conference on Human-Robot Interaction*, New York: ACM, 214-223.

Painter, Nell Irvin. 2010. *The history of white people*. New York: WW Norton & Company.

Petersen, Steve. 2011. "Designing People to Serve." In *Robot ethics: the ethical and social implications of robotics*, edited by Patrick Lin, George A Bekey and Keith Abney, 283-298. Cambridge: MIT Press.

Phillips, Elizabeth, Xuan Zhao, Daniel Ullman, and Bertram F Malle. 2018. "What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic roBOT (ABOT) Database." *2018 ACM/IEEE International Conference on Human-Robot Interaction*, Chicago: ACM, 105-113.

Powers, Aaron and Sara Kiesler. 2006. "The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes." In *Proceedings of the Conference on Human-Robot Interaction (HRI 2006)*, Salt Lake City, UT: ACM, 218–225.

Rakić, Tamara, Melanie C Steffens, and Amélie Mummendey. 2011. "Blinded by the accent! The minor role of looks in ethnic categorization." *Journal of Personality and Social Psychology* 100 (1):16-29.

Reeves, Byron, and Clifford Ivar Nass. 1998. *The media equation: How people treat computers, television, and new media like real people and places*. Stanford: CSLI Publications.

Rickford, John R, and Faye McNair-Knox. 1994. "Addressee-and topic-influenced style shift: A quantitative sociolinguistic study." In *Sociolinguistic perspectives on register*, edited by E Finegan and D Biber, 235-276. Oxford: Oxford University Press.

Rickford, John R, and Russell John Rickford. 2000. *Spoken soul: The story of black English*. New York: John Wiley & Sons Inc.

Rieder, John. 2008. *Colonialism and the emergence of science fiction*. Middletown: Wesleyan University Press.

Riek, Laurel D, Nikolaos Mavridis, Shammah Antali, Noura Darmaki, Zeeshan Ahmed, Maitha Al-Neyadi, and Amina Alketheri. 2010. "Ibn Sina steps out: Exploring arabic attitudes

toward humanoid robots." *2nd international symposium on new frontiers in human–robot interaction*, *AISB*. Leicester, 88-94.

Robertson, Jennifer. 2018. *Robo Sapiens Japanicus : Robots, Gender, Family, and the Japanese Nation*. Oakland, California: University of California Press.

Roediger, David R. 2005. *Working toward whiteness: How America's immigrants became white: The strange journey from Ellis Island to the suburbs*. New York: Basic Books.

Root, Michael. 2000. "How we divide the world." *Philosophy of Science* 67:S628-S639.

Rooth, Dan-Olof. 2010. "Automatic associations and discrimination in hiring: Real world evidence." *Labour Economics* 17 (3):523-534.

Siegel, Mikey, Cynthia Breazeal, and Michael I Norton. 2009. "Persuasive robotics: The influence of robot gender on human behavior." *2009 IEEE/RSJ international conference on intelligent robots and systems*, New York: IEEE, 2563-2568.

Smedley, Audrey, and Brian D Smedley. 2005. "Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race." *American Psychologist* 60 (1):16-26.

Smith, Sandra Susan. 2010. "Race and trust." *Annual Review of Sociology* 36:453-475.

Sparrow, Robert. 2002. “The march of the robot dogs.” *Ethics and Information Technology* 4 (4): 305-318.

Sparrow, Robert. 2016. “Robots in aged care: a dystopian future?” *AI and society* 31 (4): 445-454.

Sparrow, Robert. 2017. “Robots, rape, and representation.” *International Journal of Social Robotics* 9 (4):465-477.

Sparrow, Robert. 2019. “Do robots have race?” Manuscript under consideration for publication.

Sparrow, Robert, and Linda Sparrow. 2006. “In the hands of machines? The future of aged care.” *Minds and Machines* 16 (2): 141-161.

Stanley, Damian A, Peter Sokol-Hessner, Mahzarin R Banaji, and Elizabeth A Phelps. 2011. "Implicit race attitudes predict trustworthiness judgments and economic trust decisions." *Proceedings of the National Academy of Sciences* 108 (19):7710-7715.

Templeton, Alan R. 2013. "Biological races in humans." *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (3):262-271.

Tremain, Shelley. 2001. "On the government of disability." *Social Theory and Practice* 27 (4): 617-636.

von der Pütten, Astrid, and Nicole Krämer. 2012. "A survey on robot appearances." *7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2012*, New York: IEEE, 267-268.

Wendell, Susan. 1996. *The rejected body: Feminist philosophical reflections on disability*. New York: Routledge.

Whitehouse, Ginny. 2009. "Why diversity is an ethical issue." In *The handbook of mass media ethics*, edited by Lee Wilkins and Clifford Christians, 101-114. New York: Routledge.

Winner, Langdon. 1980. “Do Artifacts Have Politics?” *Daedalus* 109 (1): 121-136.

Womack, Ytasha. 2013. *Afrofuturism : the World of Black Sci-Fi and Fantasy Culture*. Chicago, Illinois: Lawrence Hill Books.

Zack, Naomi. 2002. *Philosophy of science and race*. New York: Routledge.

[1] In other work I have investigated the ethics of the design of robots for aged care (Sparrow, 2016; Sparrow and Sparrow, 2006), robot pets (Sparrow, 2002), and sex robots (Sparrow, 2017).

[2] There are a number of other grounds for concern about the racial politics of robots that are beyond the scope of my discussion here. In particular, critics have worried about the lack of diversity in the community that is designing and programming robots and computers. The profession of engineering is notoriously male dominated and also has a significant problem with racial diversity. We might well worry that the technologies

that are widely predicted to play a crucial role in shaping future social relations are being developed by such an unrepresentative group (Crawford 2016). Alternatively, critics have highlighted the importance of diversity (or the lack thereof) in the datasets used to train the algorithms and neural nets that will play an increasingly crucial role in the operations of robots (Howard and Borenstein 2018). Importantly, where these datasets have racial and/or gender biases the systems they inform may generate racist and/or sexist results (O'Neill 2016).

[3] Here and elsewhere in the text I use capitalised terms ("White," "Black," etc.) to refer to races and/or racial identities and lowercase terms ("white," "black," etc.) to refer to colors. One of the difficulties in writing about race is that the terms used to discuss race in any given context are themselves the products of, and inflected by, racism. My use of the terms White, Black, Brown, and Asian to refer to racial identifications should in no way be read to imply that I believe that these are biological categories.

[4] Notoriously, people often expect humanoid robots to be capable of carrying out a conversation and are disappointed when they cannot (Ishiguro and Nishio 2018, 25).

[5] The experiments carried out by Bartneck et al. (2018) 's suggested that people were responding to the surface colors of robots in ways that implied that they were attributing race to the robots without being consciously aware of this fact.

[6] The issues that might be raised by the development of sentient robots are beyond the scope of this paper. However, let me observe that it seems unlikely that sentient robots would experience racism in the same way that people do and therefore that they would evince race in the same way that people do. Equally well, though, in so far as people responded to sentient robots as if they had race, their interactions with people would still be racially inflected, which, for reasons discussed below, might well lead such robots to become aware of their own place in the local racial schema and shape their behaviour accordingly.

[7] Social constructionist accounts of gender (Butler 1990; Haslanger 2012) and of disability (Garland-Thomson 2017; Tremain 2001; Wendell 1996) have also contributed to our understanding of identity.

[8] This is not to suggest that all it would take to eliminate race is a collective act of will. As noted above, because of the history of racial identifications there are now real material differences in the social circumstances of people of different races, which would persist even if everyone were suddenly to become genuinely race blind. Nevertheless, the origins of these differences lie in the social recognition of race rather than any purported biological or metaphysical racial "essences" (Root 2000).

[9] See also: von der Pütten and Krämer (2012). A less formal, but no less compelling, measure of the colour of robots can be achieved by performing a Google Image search for "humanoid robot" or "android": such a search returns page after page of images of white robots.

[10] Because I'm interested in the design choices of engineers and the *surface* colours of robots, these descriptions neglect the colour of a robot's actuators, which are often grey or black, except where these are so prominent as to clearly controvert the account of the colour of the robot provided here. It is also worth noting that several of the silver or brightly coloured robots have faces that are white. Five of the robots with white skin tones have facial features and/or hair that make it plausible to believe that they should be understood as Asian.

[11] See: https://www.softbankrobotics.com/emea/en/pepper; https://www.softbankrobotics.com/emea/en/nao. Remarkably, for all of the press about humanoid social robots, these are, to my knowledge, the only two humanoid social robots capable of reasonably sophisticated social interactions that have been sold in any number to anyone other than robotics researchers.

[12] Although, as Louis Chude-Sokei (2016) has argued, and I will discuss below, the historical race politics of robots is significantly more complex than I allow here. In particular, the history of cultural representations of robots in film and literature means that, despite the gleaming white surfaces of contemporary robots, their Whiteness is somewhat precarious. Nevertheless, the (admittedly limited) empirical evidence we have suggests that people do identify contemporary white robots as White.

[13] This is not to deny that, once race categories exist, skin colour may be relevant to their application. However, the nature and content of these categories is not determined by the reflective properties of skin: neither can the question of whether a particular individual is White be settled simply by their skin colour.

[14] Insofar, as much science fiction has also been driven by unconscious – or, indeed, explicit – fears about the loss of White privilege associated with the end of slavery and with decolonisation, people of colour have, in an important sense, remained at the very centre of representations of the future (Chude-Sokei 2016: 83-127; Rieder 2008). Nevertheless, in the mainstream of science fiction, even when playing a central role, people of colour have typically been portrayed as confined to the margins of a society imagined as White and been represented either as a threat to "civilisation" or as the nature which technology seeks to control – or both. For a collection of discussions that focus on the representations of people of colour in science fiction, see Lavender (2014).

[15] For a discussion of the role played by different communities in settling the question of the representational content of robots, see Sparrow (2017).

[16] It is remarkable that neither Bryson's (2010) paper, "Robots should be slaves," nor Petersen's (2011) book chapter, "Designing People to Serve," discusses the real world history of slavery.

[17] Although he does not mention slavery, Levy (2009) suggests that the likelihood that our treatment of robots will shape our treatment of other human beings gives us reason not to mistreat robots.

[18] I have discussed the extent to which it is plausible to think that our interactions with robots will shape our behaviour towards other people, as well as expressive content of such interactions, at length in Sparrow (2017).

[19] It is not too much of a stretch to suggest that, culturally speaking, contemporary robots are, implicitly, Black souls in White bodies. In conversation, Aaron Steinfeld suggested to me that one might therefore think of existing robots as enjoying "borrowed whiteness" as is, some have suggested, the case with Jews in America.

[20] Of course, in some contexts engineers are already aware of this fact and deliberately try to make robots less human in appearance in order to reduce the expectation amongst users that the robots will be able to do the things that humans can do. What is novel about the argument I have developed here is that it establishes that there may be reasons to make robots appear other than human even when they *do* become capable of a level of performance that meets or even exceeds what people expect of them. For a different argument to the same conclusion in another context, see Sparrow (2017).