Why Robots Can’t Become Racist, and Why Humans Can

MATTHEW T. NOWACHEK

Of all the areas of research within philosophy, few have proven as broad in interdisciplinary reach as race theory. Unlike some more traditional branches in philosophy that tend not to venture too far beyond the walls of philosophy departments, race theory has from the beginning developed in conjunction with other disciplines as varied as biology, anthropology, sociology, linguistics, and economics. The outcome of this approach to research has been and continues to be a fruitful and enlightening discourse on the subtleties and complexities of race and racism. In this essay I would like to proceed in the same interdisciplinary spirit by drawing into a discussion of racism several insights from phenomenology as well as studies in robotic artificial intelligence (hereafter AI). Although a precedent has been set for use of the former in race discourse and thus such use has achieved a certain amount of respectability (see Gordon; Alcoff 179-194; Yancy 1-32), the latter has received little to no attention and therefore may strike the reader as either odd or irrelevant. Nevertheless, this dual emphasis is intended for a particular purpose, namely to help tease out in an interesting manner the deep and complex interrelation of racism, ontology, and learning.

More specifically, this essay engages in large part with the question of whether robots can become racist in the manner humans can and what it would require for robots to do so. It is important to note that the emphasis here is not on whether or not robots can be racist. Clearly, robots can be racist in the same sense that artifacts such as books and bumper stickers or
institutions such as universities and banks can be racist, that is, as products or projections of human dispositions, behaviors, and intentions. However, to talk about the more interesting question of robots becoming racist is to shift the discussion towards the classic topic of the possibility of learning. How one becomes racist is a function of how one learns racism as well as a function of the type of being which learns, and as such we cannot talk about learning without also talking about ontology. Thus, counter to the views of certain scholars (see, e.g., Lycan), the question of robots learning racism in the manner of humans is not strictly empirical, but is also deeply philosophical. In addition, the phrase “in the manner of humans” is not insignificant, for regardless of what we may say concerning robots, humans do appear to experience and to pass on racism in a very real way. Therefore, I suggest that in order to better understand this complex human phenomenon a running comparison between robots and humans can prove useful. Although this comparison most certainly contributes to the general critique of the AI research program in that it reveals yet another crucial difference between robots and humans, I intend the comparison in this essay primarily as a heuristic device that is directed, on the one hand, towards clarifying and criticizing what I identify as a certain Cartesian view of what racism is and how it is learned, and, on the other hand, towards supporting an alternative phenomenological view. In drawing such a comparison, I thus side with Ian Barbour when he rightly notes, “Our view of computers and robots…will influence our own self-understanding” (89).

My thesis is fairly elaborate due to its interdisciplinary nature, but it can nevertheless be summarized in the following manner. With respect to the guiding question of the essay, I argue that robots cannot become racist insofar as their ontology does not allow for an adequate relation to the social world which is necessary for learning racism, where racism is understood in terms of a social practice. This is revealed most clearly in the failure of robots to manage common-
sense knowledge in its tacit and social forms—a problem that has come to be known as the common-sense knowledge problem (Dreyfus, “Why Heideggerian AI Failed” 248). In contrast, I argue that humans can become racist precisely because human ontology, characterized by the embodied, porous, and shifting nature of the human self with and in relation to its social background, allows for racism to be learned through mediums such as habit and skill acquisition. As such, the human self is able to cope with the common-sense knowledge that proves so challenging for AI. In short, then, the specific reasons underlying why robots are unable to become racist can prove informative both for critiquing a flawed view of how racism is learned and for helping to illuminate precisely what it is that makes it possible for human beings to become racist.

To argue for this thesis, I proceed for the remainder of this essay in three sections. First, I begin by discussing the conceptual and ontological status of race and the implications this carries for racism and how racism is learned. Second, I consider the ontology of robots in light of the interesting debate surrounding the development of AI. Here I draw heavily on the critical phenomenological work of Hubert Dreyfus to outline what I see as important challenges facing robots with respect to learning racism. Third, to continue the discussion from section I, I turn to the phenomenology of Maurice Merleau-Ponty to suggest a way in which human ontology, in contrast to that of robots, makes possible the learning of racism through habit and skill acquisition. We may begin with race and racism.
I. Race, racism, and the significance of the social context for learning

Before turning to the ontology of robots and humans, we will first need to discuss what we mean by the terms “race” and “racism,” as well as how these terms are related. This conceptual work is important for our project insofar as what we take racism to be and how we conceive of learning such racism is significantly shaped by the ontology of race. (We turn to the ontological question of what sort of being the potential learner is in sections II and III.) In what follows, I argue two main points. First, we have good reason to think race is not an objective natural kind, but rather objectively real in the sense of a social kind. Therefore, second, racism, as conceptually tied to race, cannot be grounded merely on “objective facts of the world,” but must turn to the social context to gain its meaning. Based on these two points, I conclude this section by sketching three foundational criteria required for learning racism.

a) The ontology of race

Early attempts to explain race have done so in terms of a reality built into the structure of nature. Within this view, known as race naturalism (Haslanger), races are seen as natural kinds or biological essences that are objectively the case independent of human classification, and by which it is possible to give a clear and definite taxonomy of human beings. Despite the historical dominance of this view, it has run into significant difficulty, particularly in recent years. The chief problem, which has been made especially acute by the completion of the Human Genome Project and its conclusion that no such thing as a race gene actually exists, is the inability of race naturalists to provide an ostensible grounding for race as a natural biological kind (Dupré). This failure has in turn led thinkers such as Kwame Anthony Appiah and Naomi Zack to adopt a race eliminativist stance by which, on moral grounds, they wholly reject the concept of race (see Appiah; Zack).
This does not mean, however, that the project of connecting race to biology has been entirely abandoned. In response to the critique of the robust realism of previous generations, contemporary race naturalists have shied away from describing groups in terms of racial essences. Nevertheless, they still attempt to retain race as a pragmatic tool for addressing certain biological issues such as “racially” linked health problems (see Lee et al.; Risch et al.). It is important to note, however, that even though such theorists still characterize race with respect to biology, what is meant by race is not a strictly objective natural kind, but rather something more along the lines of a statistically prevalent cluster of genes shared by a group of individuals who also have in common a specific physical malady. As such, the lines determining who is and who is not of a certain “race” are not drawn absolutely, but rather they are largely dictated by the conventional parameters established by researchers through the questions they pose in their research projects.

The crucial conclusion to draw here is that pointing to biological factors is not enough to formulate a conception of race that proves non-arbitrary. Although biologists are able to identify a single race, namely the human race, in the absence of necessary and sufficient conditions for racial distinctions, the most biologists can offer concerning races is a complex spectrum of human diversity. In other words, once the anchor of realist biology has been relinquished, any attempt to derive an account of race from morphology (in the case of early race naturalism) or other biological factors (in the case of more recent race naturalism) can at most expect a vague concept.

One attempt to deal with this vagueness while still retaining the biological connection to race is to describe race in terms of something like family resemblances, or what Jorge Gracia calls “the genetic common-bundle view” (83-84). As Gracia argues,
part or what ties members of a race is the same kind of thing that ties some members of a family. There may be no common features to all of them, but nonetheless they belong to the same group of persons because they are genetically related...Moreover, they share in a pool of features that need not be common to all the members of the group. It also means...that any requirements of coherence and homogeneity do not apply. Racial groups are not coherent wholes composed of homogeneous elements. Like families, races are related groups of human persons with different, and sometimes incompatible, features. Homogeneity is not one of their necessary conditions. (88)

In the same way that for Ludwig Wittgenstein language holds together not with respect to a central core, but rather through “a complicated network of similarities, overlapping and criss-crossing” (Wittgenstein 66), for Gracia race as a concept lacking an essential core holds together as a diverse and shifting interrelation of differing physical features. The important question to ask at this point, however, is whether this genetic family approach alone can provide us with a satisfactory account of race. It appears that the answer is no, for in the same fashion as with contemporary race naturalism, as soon as we divorce our biological concept of race from a realist framework, we are still left with the problem of how to identify (since nature does not) which genetic features are relevant for recognizing certain family resemblances. To answer this type of dilemma with respect to language, Wittgenstein looks to the social context and the social practices within which language develops as the basis for determining relevancy. For Gracia a similar turn to the social is necessary if we are still to derive a non-arbitrary unity out of a vague set of diverse features. In short, to be a coherent concept, race must look to and depend upon the social context.

By pointing to the crucial role of the social realm for deriving an account of race, we are not thereby committed to thinking of race as a relative or empty concept. In contrast, for theorists within the social constructionist framework, even if race is entirely a social construct it can still possess objective reality. As Charles Mills puts it, “there is conceptual room for a view of race as
both real and unreal, not ‘realist’ but still objectivist’ (47). This is an important move because it offers us a mechanism for judging what is or is not characteristic of a certain race in a manner that simultaneously takes into account the depth by which race permeates lived experience. Again, to quote Mills,

race is a *contingently* deep reality that structures our particular social universe, having a socially objective and causal significance that arise[s] out of our particular history…Because people come to think of themselves as “raced,” as black and white, for example, these categories, which correspond to no natural kinds, attain a social reality. Intersubjectivity creates a certain kind of objectivity. (48)

In sum, with respect to the ontology of race we have provided good reason to reject the characterization of race as a natural kind, but we have also made a case for embracing race as a social reality. Thus, unlike the race eliminativists, we need not dismiss race altogether if we have the good alternative of redefining the ontological categories underlying it, namely by moving from a realist ontology to a social ontology. With this brief outline completed, let us turn now to racism.

*b) A social account of racism*

In light of the overwhelming *prima facie* evidence we have for the reality of racism, we may assume its existence as our starting point and move backwards in order to fill out in more detail how it functions. As such, we will then be able to understand better what we mean when we talk about learning racism. Although, as Paul Taylor rightly acknowledges, racism is notoriously difficult to tie down and therefore any definition runs the risk of oversimplification (32-33), we may still offer three important points for a working conception that we will employ in later sections.

First, in order to make sense of racism, we cannot divorce it from the ontology of race. This does not mean that every racist possesses a philosophically developed theory of race. (The
opposite is more likely the case.) Rather, it means that racism is conceptually parasitic upon some account of race regardless of how clearly this account is understood or even if it is explicitly recognized. To put it differently, in order for racism to be racism, it must be juxtaposed with or dialectically directed against race in some real fashion. If this were not the case, we would have a difficult time making the distinctions we do between racism and other forms of oppression such as sexism or classism.

Second, because racism is conceptually dependent upon race, and since race is only fully intelligible as a social concept, the only way to make sense of the phenomenon of racism must likewise be in terms of the social context. However, just as with the concept of race, this need not entail that we think of racism as any less real, but rather in order for racism to be real means that it must take its reality from the social context in which it is practiced. In his own account of racism as “disregard” Taylor captures this social aspect of racism well by describing it as “the withholding of respect, concern, goodwill, or care from members of a race” (33). As this definition displays, racism operates as a type of value judgment concerning a person or a group of people. Nevertheless, this value judgment is one that cannot be objective in any realist sense (since race is no natural biological kind), but rather it is only meaningful within a network of social relationships and practices in which values are socially constructed and assigned. Thus, to disregard persons of a certain race, whether this is done intentionally or unintentionally, is to disregard these persons in a specifically social manner.

Third, because racism operates as a social practice, it becomes (or is) embedded within societal norms in a manner that for the most part goes unrecognized. Some theorists such as Shannon Sullivan have displayed how racism takes the form of unconscious habits of white privilege (Sullivan). Others, like Roxana Ng, who is herself drawing on Antonio Gramsci’s
account of normalized ideology (Ng), point out that in becoming the societal norm, racism becomes buried within language itself. For Ng, this entails that our racist dispositions and behaviors operate as a form of (perverted) common-sense, such as in the “common-sense” statement, “Blacks are good at sports but not at academic subjects…” (194). In other words, because racism operates within a society as a normalized standard, it more often than not is expressed merely in terms of “what everyone simply knows.” As a product of a long and historical process of habituation and conformity to “whiteness as the transcendental norm” (Yancy 3), racism which functions as common-sense is one of the most powerful and damaging expressions of the phenomenon. In addition, as we will see in section II, due to the specific ontology of robots, this aspect of racism presents the greatest hindrance to robots becoming racist.

c) Significance of the social context for learning racism: Three criteria

To conclude this section, we may now draw out the implications of the discussion so far in order to formulate three foundational criteria for what is required if one is to learn racism. We may think of this as a kind of basic racism litmus test. To summarize what we have argued above, race is no biological real kind, but rather a social kind, and because of the conceptual dependency of racism upon race, racism is best understood as originating and operating within a social context as a common-sense type of social practice. Because of this specific social nature of racism, our criteria for what is required in order to learn racism must themselves reflect the significance of the social world. The following is in no way meant as an exhaustive list, but we may nevertheless say that for any being to become racist in the social manner we outlined above would require the satisfaction of at least three interrelated criteria:
1) *Social Receptivity*: a being must be able to receive in a cognitive or non-cognitive fashion racial cues from a social context about a specific individual or a group.

2) *Social Interpretation*: a being must be able to interpret in a cognitive or non-cognitive fashion which of these racial cues are relevant to a specific situation and which racial cues apply to a specific individual or a group.

3) *Social Application*: a being must be able to apply in a cognitive or non-cognitive fashion these interpreted racial cues in a discriminatory fashion against a specific individual or a group.

To illustrate this, let us use a simple example. On the one hand, we may say that stones cannot fulfil the first criterion (not to mention the second and third), and therefore they cannot become racist. On the other hand, since humans apparently can fulfil all three (as we will return to later in section III), we have good reason to think that for them racism is a real possibility. The case of robots is interesting precisely because it falls somewhere between these two extremes. Although I argue in the next section that robots cannot ultimately fulfil these criteria and therefore cannot become racist, the reasons I provide for this reveals in an interesting manner why and how human beings can.

**II. Robots, ontology, and learning, or why robots can’t become racist**

Given our conclusions from the previous section, what can now be said concerning the topic of robots and racism? The main claim I would like to develop is that robots cannot become racist insofar as their ontology prevents them from coping with common-sense knowledge in its tacit and social forms, something which is required to learn racism and to fulfil the three criteria outlined above in section I. To support this, I locate my argument within the broader debate between AI and the critical work of Dreyfus. With this argument I am not attempting to set up AI as a straw-person, but rather to show how the inability of robots to become racist can teach us a
great deal (at least in a negative fashion) about racism as a human phenomenon. More specifically, the inability of robots to become racist due to the dependence of the AI research program upon a certain form of Cartesian ontology provides additional support for the phenomenological description of how humans learn racism, which we develop in section III.

a) Development and difficulties in AI

From its inception, the goal of AI has been to produce an artificial form of intelligence in robots that can mirror if not exceed human capabilities. AI in its current form is the product of two major movements to attempt this project, namely classical symbolic-representational AI, or rather “GOFAI: Good Old-Fashioned AI” (Haugland 112) and what has recently come to be known as embodied “Heideggerian AI” (Wheeler 223; Dreyfus, “Why Heideggerian AI Failed” 253). Although the former is currently a degenerate research program (Dreyfus, What Computers Still Can’t Do ix-x), the central problems that have plagued GOFAI continue to prove problematic for AI in its more recent forms.

In its early development, AI worked with certain ontological assumptions about the world, robots, and human beings, which can be best characterized by the physical symbol system hypothesis (Newell and Simon). According to this hypothesis, a physical symbol system “has the necessary and sufficient means for general intelligent action” where such action means “any real situation behavior appropr[i]ate to the ends of the system and adaptive to the demands of the environment…” (Newell and Simon 116). In this view, the world is seen as a collection of data, and intelligence (both human and robotic) is thereby understood as the processes of data collection, data processing, and data distribution. Consequently, learning is described as an operation of formal rule-based symbol manipulation. For a generation of researchers, this view became the guiding basis for AI projects and the source of a growing optimism about the
promises of AI technologies. Nevertheless, after only a few years of operation the GOFAI research program came to a grinding halt due to its own ontological limitations and the increasing philosophical criticism of it.

At the heart of the difficulties with GOFAI as well as with contemporary forms of AI, is the problem of common-sense knowledge, or we could also say the problem of relevancy. Simply put, this problem centers on the difficulty that information-processers face in drawing out relevant facts within a fluid world and separating these from irrelevant facts so as to act within a given situation in an appropriate manner. Dreyfus, in his broad criticisms of AI from the past four decades, assumes this as one of his central points of contention with the overall AI tradition. To illustrate this specific challenge, Dreyfus draws on the relatively simple activity of horse-betting. For any given race, the number of possible relevant factors for placing a successful bet is indefinitely large, and for this reason a good bet requires a healthy dose of handicapper’s know-how. As Dreyfus notes,

In placing a bet we can usually restrict ourselves to such facts as the horse’s age, jockey, past performance, and competition…but there are always other factors such as whether the horse is allergic to goldenrod or whether the jockey has just had a fight with the owner, which may in some cases be decisive. Human handicappers are no more omniscient than machines, but they are capable of recognizing the relevance of such facts if they come across them. The artificial intelligence approach to this human ability would have to be to give the machine knowledge about veterinary medicine, how people behave when they fight their employers, and so forth. But then the problem arises of sorting through this vast storehouse of data. To which the answer is that all this information would be properly coded and tagged in the machine memory so that the machine would just have to do a scan for “horse-race betting” and get out the relevant material. But not all relevant material would have been encoded with a reference to this particular use. (What Computers Still Can’t Do 258)

One clear difficulty this example points to is the issue of a regress of rules for determining what is relevant in a given context. For human beings, this issue is not normally problematic in that common-sense know-how itself acts as the terminus for the regress. To use an example from
Yehoshua Bar-Hillel, most of us who speak English with some basic proficiency simply know that when we hear the statement, “Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy” (74), we automatically take into consideration all the relevant information that we need in order to grasp the meaning and significance of the statement, such as the possible varied meanings of the word “pen” as well as the common size of ink writing utensils. Although GOFAI is effective with highly formalized tasks such as running complicated algorithms, in contrast to humans it is unable to manage the sort of knowledge that characterizes everyday experience because it cannot sort through the built-in ambiguity of situations in order to discern what is relevant for any specific situation. As Dreyfus concludes, without a viable alternative to formal rules for determining relevancy, all AI can turn to is rules for rules for rules ad infinitum. Thus, the everyday world remains far too challenging for robots to function properly within—a fact that for Dreyfus proves true not only of GOFAI but also of the more recent Heideggerian AI.

b) Limitations in robot ontology

The failure of AI in the area of common-sense knowledge can be traced to deeper ontological limitations in robots that extend from what we may identify as a latent form of Cartesianism. As we have suggested above, the robot addresses the world not from a position intimately situated within the world, but rather in the role of data processor set apart from the world as a disconnected observer. What this means is that the robot takes in information through various means and attempts to order this information all the while missing the porous relationship between subject and world that is necessary to acquire common-sense knowledge. For Dreyfus, drawing on the thought of Martin Heidegger, the crucial ontological point concerning learning is that it must occur as a function of being-in-the-world:
Heidegger’s important insight is not that, when we solve problems, we sometimes make use of representational equipment outside of our bodies, but that being-in-the-world is more basic than thinking and solving problems; it is not representational at all. That is, when we are coping at our best, we are drawn in by solicitations and respond directly to them, so that the distinction between us and our equipment vanishes. (“Why Heideggerian AI Failed” 254)

To put Dreyfus’ point another way, the only manner by which a robot could successfully cope with common-sense knowledge and gain access to the social world, would be through an embodied medium that is already woven into the structure of the world. However, with its strict subject-object (Cartesian) ontology, AI has failed to come anywhere close to emulating this, and for Dreyfus such a failure is a tell-tale sign of a larger problem with the entire AI project.

More recent AI researchers have taken this ontological point seriously and have started to develop conceptual frameworks that focus much more on the robot as situated within its environment (see Brooks and Steels; Brooks; Wheeler; Pfeifer and Bongard). Such so-called Heideggerian AI is an improvement over GOFAI in one sense in that it takes greater account of the world itself as the model for action. Yet, one major shortcoming in this embodied form of AI is that it still understands the robot in representational terms as a subject engaging the object of the world in order to solve problems. In this fashion, although the same Cartesian subject-object framework is pushed outward a level into the world, the underlying ontology remains unchanged. For Dreyfus, this is still a failure to understand the way in which human existence at its most natural and basic state is not the cognitive act of solving problems at all, but rather an unreflective non-analytic coping within and through the background. As he notes, it “is not that the mind is sometimes extended into the world but rather that, in our most basic way of being [as humans]…we are not minds at all but one with the world” (“Why Heideggerian AI Failed” 255).
In short, because AI has not abandoned its underlying Cartesian ontology, even an embodied robot is still solipsistically disconnected from the world as something of an outsider.

c) Robots and learning racism

By now the connection between our discussion of AI and our question about robots learning racism should be apparent. We can thus finish this section by returning briefly to our racism litmus test from section I. In light of our conclusions above, how do robots measure up?

With respect to the criterion of social receptivity, we have good reason to think robots cannot fulfil it. Because “facts” about race cannot be objectively based in a biological essence, racial cues cannot be delivered and received through any objective measure such as DNA testing. For this reason, the robot’s ontology must support receptivity to cues in the form of social relations. As we pointed out above, however, the ontological structure of robots as solipsistic and atomistic problem-solvers entails that robots are good with discrete concrete information, but they are neither good with common-sense social knowledge nor are they able to tolerate ambiguity. Because of the specific ontology of race as biologically vague yet socially concrete, racial cues informing racism remain entirely ambiguous unless mediated through social norms that give them meaning and value. However, since the ontology of robots draws a sharp distinction between the robot and the social world in such a way that makes common-sense access to these norms unattainable, racial cues remain ambiguous for robots. As such, robots have no meaningful way to receive the racial cues necessary to become racist.

The problem of receptivity is closely related to the problem robots face in fulfilling the criterion of social interpretation. If we grant for the sake of argument that robots could in some manner receive some social cues about race, we are still faced with the massive problem of how, in a non-static social environment, robots are to determine which cues are relevant for which
people, in which contexts, and how these are to be applied in an “appropriately” racist manner. The striking thing about racism as a social phenomenon is that it is never entirely uniform, but rather it is always shifting in a fluid fashion with respect to the social context. For example, racism in America directed against people of Middle-Eastern descent functioned in a very different way prior to the 9/11 terrorist attacks than it did in the days and weeks after. To explain this one cannot turn to a strict formula, but rather one must take into account a complicated network of stereotypes, emotions, nationalistic sentiments, historical narratives, religious identities, and so forth. Again, human beings are able to negotiate this complexity through appeal to social common-sense; however, as we argued above, because of their latent Cartesian ontology, this form of knowledge is precisely that to which what robots do not have access. As such, even if robots could receive racial cues, the problem of relevancy still undermines the ability of robots to carry out any meaningful social interpretation of these cues.

Finally, with respect to the third criterion of social application, robots face yet another difficulty. The problem for robots is not necessarily in acting certain ways that may be perceived as racist by others—something of a racism Turing test—for such a perceived racist action could in actuality be nothing more than the result of a random generator which happened to coincide with the racist social norm within the social context. This random action, however, would not satisfy our requirement of acting in a discriminatory fashion against a specific individual or group. In order for the robot to act in this fashion, it would need to fulfil the first two criteria so as to receive and to interpret racialized cues in such a way as to apply them “appropriately” in a novel fashion against people within a new and unfolding situation. As we have already adequately shown, however, the ontology of robots presents an impassable hurdle to success in this area.
To conclude, we may note that because of the specific ontology with which robots operate, an ontology that we have identified as Cartesian in nature, robots are unable to fulfil any of our three criteria for learning racism. Thus, we may say that since robots cannot learn racism (neither consciously nor unconsciously), we cannot talk in any meaningful sense about robots becoming racist. With respect to race theory in general, the implications of this conclusion are significant. The inability of robots to learn racism entails that any view concerning racism which operates with a similar underlying Cartesian ontology, namely one in which the subject is seen as existing distinct from the world except through the cognitive relation of knower to known, must also be rejected. Not only does this challenge the all too common view that racism is merely a cognitive problem of ignorance or false beliefs, but it also forces us to take seriously the role that human embodiment plays in learning racism. In order to develop this latter point, we may now turn to our final section.

III. Humans, ontology, and learning, or why humans can become racist

In light of our conclusions above, what may we now say about the topic of humans and racism? Up to this point in the essay, I have argued that robots are unable to become racist while at the same time I have assumed that racism operates as a real social phenomenon experienced and practiced by human beings. This discrepancy, however, is not merely quantitative, as if better development in AI would bridge the gap. Rather, the difference is fundamentally qualitative. In support of this claim, we may draw on Merleau-Ponty’s phenomenological account of the embodied self as well as how this self interacts with and through the world in which it is situated.
a) Merleau-Ponty’s ontology: Embodiment, habit, and skill acquisition

In his groundbreaking book, *Phenomenology of Perception*, Merleau-Ponty develops his phenomenological and ontological insights in response in part to the reigning Cartesian ideology of his day. To do so, he focuses on the centrality of embodiment for human existence. Embodiment for Merleau-Ponty means that the self is fundamentally connected to the world in such a manner that sharp distinctions between self-other, subject-object, and inside-outside no longer hold. In this view, the background is treated not merely as a framework on which the human person exists as a distinct knower, but rather it is what the human being is inextricably and inseparably situated within and communicating with—what Merleau-Ponty identifies as the “intentional arc” (157). To illustrate what he means with embodiment and to argue that this provides a qualitatively different ontology than that of Descartes, Merleau-Ponty gives significant attention to the human capacities of habit and skill acquisition. These prove crucial for our discussion of how humans learn racism in contrast to robots, and therefore we may give a brief account of each.

To understand what Merleau-Ponty means by habit, it is important to see how his conception of habit relates to his account of consciousness. For Merleau-Ponty, consciousness is not a distinct mental act or an ideal representation as it is for Descartes, but rather it functions as a “being-towards-the-thing through the intermediary of the body” (160). What this means is that the mental is given no priority over the body in the act of consciousness, but rather “consciousness projects itself into a physical world and has a body, as it projects itself into a cultural world and has its habits” (158). Central to this projection of consciousness is motility, or rather the power of motion. Again, in contrast to Descartes, Merleau-Ponty argues that bodily movements are not the products of mental representations, but rather it is through the process of
moving and acting within the world that consciousness is made possible. As he notes, “Consciousness is in the first place not a manner of ‘I think that’ but of ‘I can’” (159).

Habits, therefore, begin to take form out of the physical activity of existing through motion. In this manner, habit is not the result of a process of pure cognition or formal rule-application, but rather it develops through the corporeal schema that is already at work within the world. Thus, although habit operates as a form of learning, it is not learning in any intellectualist sense, but rather learning as a product of the body’s physical presence within the world, which already carries the “stamp of movement” upon it (165). Just as a dancer is able to learn a new dance step insofar as his or her body is already receptive to and incorporating the step through bodily motion, the human being is able to form habits through what Mary Barral calls “a sort of intelligence of the body” (137). For Merleau-Ponty, this forces us to reconsider our conception of understanding:

the phenomenon of habit is just what prompts us to revise our notion of “understand” and our notion of the body. To understand is to experience harmony between what we aim at and what is given, between the intention and the performance—and the body is our anchorage in a world. (167)

To re-conceptualize the understanding in this manner is not to abandon significance, but to seek significance in a new habitual form: “The acquisition of a habit is indeed the grasping of a significance, but it is the motor grasping of a motor significance” (165). Thus, through his account of habit as motility, Merleau-Ponty gives a new meaning to the word meaning, and thereby redefines how we understand significance and learning in relation to the world.

To parse out in a more concrete manner what he means by habit, Merleau-Ponty provides several examples. Consider the woman who places a feather in her hat. At first, the woman may find the addition to her hat to be awkward, but after a brief period of maneuvering with the
feather, the woman has learned, without mental calculation, the most proficient way to comport her body so as to function normally without damaging the feather (such as when she walks through a doorway). As Merleau-Ponty notes, “She feels where the feather is just as we feel where our hand is” (165). Other examples outlined in the Phenomenology of Perception where human beings become habituated to their surroundings include the novice learning how to drive a car, the blind man morphing into his guiding stick, or the organist who becomes quickly comfortable playing on an unfamiliar organ. The central idea behind these examples is that habit forms as a type of adaptation of the body to physical surroundings that is only possible if the strict distinction between self and world is relinquished.

As soon as we begin to talk about habit not simply in negative terms as a bracketing of sensory stimuli such as a foul odor, but rather in the positive terms of development and proficiency, we are already moving into the notion of skill acquisition. For Merleau-Ponty, skill is possible because habits are never static, but are always being refined through practice. On the one hand, practice may take the form of an intentional behavior such as in the case of a potter who intentionally sits at the potter’s wheel throwing pots for eight hours every day. Yet, on the other hand, because habit is embodied, practice may also be carried out unintentionally such as in the case of infant language acquisition. Apart from any conscious recognition, a child may become increasingly skilled at communicating in a language simply from existing within a social context and conforming to that context’s linguistic norm. To put it in a different way, the former sense of skill acquisition represents learning as directed behavior, whereas the latter sense of skill acquisition represents learning as adaptation. What is most important to recognize in both cases, however, is that any development in skill requires a movement away from cognition into deeper embodied and unreflective engagement with the world. To quote Dreyfus,
What emerges from this model of human skill acquisition is a progression from the analytic, rule-governed behavior of a detached subject who consciously breaks down his environment into recognizable elements, to the skilled behavior of an involved subject based on an accumulation of concrete experiences and the unconscious recognition of new situations as similar to remembered ones. ("Misrepresenting Human Intelligence" 384)

Skill as a habitual exercise cannot be reduced to the mind or to the body alone, but rather it flows from an intertwining between self and world as a mediating alternative that places, as Merleau-Ponty nicely describes it, “knowledge in the hands”: “If habit is neither a form of knowledge nor an involuntary action, what then is it? It is knowledge in the hands, which is forthcoming only when bodily effort is made, and cannot be formulated in detachment from effort” (166).

b) Humans and learning racism

With this ontological account of embodiment expressed through habit and skill acquisition as our foundation, we are now able to argue for a way to understand, in contrast to robots, how human beings can become racist. As we pointed about above, the embodied human is not an atomistic being distinct from the world and neither is the embodied human distinct from the social context in which it is situated. Rather, the ability to learn and adapt to the world through habit and skill acquisition displays how the human self is porous and shifting in relation to its background. Nevertheless, despite the indispensible positive and necessary role habit and skill play for helping humans “be at home in” (Gordon and Tamari 72-74) and cope with the world, it is precisely these capacities which unfortunately also make the learning of racism possible.

Consider first the relation of habit to common-sense knowledge in learning racism. As Sullivan rightly notes, habits are “dispositions for transacting with the world, and they make up the very beings that humans are” (2). What this means for racism is that through specific physical habits that a human being develops within a cultural context he or she is able to tap into the common-sense knowledge of that context in a manner that bypasses any difficulties with
determining the relevancy of racial factors. To put it differently, one’s habits simply reflect the cultural racist norm in a fashion that is already “appropriate” and “relevant.” As an example of this, consider George Yancy’s description of his experience in an elevator with a white woman. Upon Yancy’s entrance into the elevator, the woman’s body has already made the relevant racist assessment of Yancy (a black man) through the manner by which it (and she) reacts:

the white woman’s body shifts nervously and her heart beats more quickly as she catches a glimpse of my Black body. Anxious, her palms become sweaty. There is an acute sense of panic. There is the difficulty swallowing, dry mouth, nausea. All of this happens at the level of the somatic. In other words, this blindness is also embedded within her encrusted bodily ways of engaging-in-the-world; it is a form of blindness that involves her bodily ways of traversing spaces within which Black bodies are present. Moreover, her constructed ignorance is cultural, and not just an individual act. She does not realize the subtle, habitual performances that she enacts in order to sustain the socially constructed nature of her gaze, and, hence, to continue to perpetuate the distortion of my Black body as criminal. (22)

In contrast to robots, the woman has no issues with determining relevancy of racial cues because in this situation when she is alone with “this type of man” her habitual relation to the world serves as her medium for adopting common-sense knowledge in the form of racism. In other words, her body has already made the culturally relevant judgment for her. This is because her racism is not restricted to “cognitively false beliefs,” but it already also saturates her embodied habitual contact with the world (22).

In addition, Merleau-Ponty’s account of skill acquisition provides us with another insightful way to understand how humans are able to learn racism. In a context such as American society that is structured around white privilege and white normativity, conceptions of whiteness become associated with the values of “good,” “beautiful,” “successful,” and so forth. As such, the practices and habits of the culture thereby also lean toward this standard. As Sullivan notes, “In a world filled with white privilege, habits that privilege whiteness will result, and these
habits in turn will tend to reinforce the social, political, economic, and other privileges that white people have” (4). Insofar as conformity to a racist norm represents “success” in a culture, skill acquisition in terms of the skill of conforming to the standard becomes a crucial way in which racism is learned by human beings.

Unlike robots which remain foreign to the social context in that they are unable to receive and to interpret racial cues in a way that would allow them to adapt appropriately, human beings may be foreigners to a specific social context, but because of their embodied existence within the world they are never entirely foreign. Rather, humans are able to overcome foreignness through skill. For example, as Noel Ignatiev illustrates, in the decades following their immigration into America, the Irish underwent a successful process of “becoming white” largely by acting out more oppressive and brutal forms of racism against blacks than was practiced even by the white Anglo American power structure (Ignatiev). Such successful conformity can certainly be considered a type of skill insofar as it served as the (inappropriate) basis for Irish cultural adaptation. However, interestingly this form of skill acquisition is not restricted to immigrants, but also has an effect even upon oppressed groups, as is often seen in the attempt of non-whites to “whiten” themselves (Davis), or in black children’s overt preference for whiteness (Clark and Clark). What is important to note about all of these cases is that racism need not necessarily function as a chosen disposition, but more often than not it is coexistent with an unconscious skillful conformity to a racist norm that is passed through embodied mediums as seemingly banal as language, gestures, eye contact, or body positioning. In short, because of this embodied ability to acquire skill, humans are therefore able to learn racism.

In bringing this section and this essay to a close, we may draw together our argument by turning one more time to our racism litmus test from section I. With respect to social receptivity,
humans are able to receive racial cues in a cognitive manner, but more importantly than this, humans are also able to pick up racial cues from a social context through habit and skill acquisition. This is possible precisely because the human embodied self, unlike the robot, is already situated within as well as “transacting” with the world (Sullivan 23-25, 82, 148, 163-4). Similarly, humans are able to interpret which racial cues are relevant to which situation and to which people because, as is evident with Yancy’s elevator effect, their embodied relation to the world already acts as the link to the racist common-sense knowledge that is embedded within the racist norm of the culture. And third, because humans are able to receive and interpret racial cues, they are also able to apply these same cues in a discriminatory fashion that is “appropriate” to the racist norm in which they are situated. Thus, we can conclude that because of the specific ontology with which humans operate, humans are able to fulfil all three criteria for learning racism. As such, we may therefore say that since humans can learn racism, we can talk in a meaningful sense about humans becoming racist in a distinctly human manner—a manner that simply does not apply to robots. Understanding this difference not only provides us with an insightful glimpse into the nature of racism, but it also opens up another potential avenue by which racism can be understood and thereby confronted. In displaying the inability of robots to become racist we have good reason to dismiss the Cartesian view of racism that sees the phenomenon as largely or strictly a cognitive problem of ignorance or false beliefs that arises in the asymmetrical relation of a subject to the world. Additionally, our heuristic device should encourage us to turn our attention to exploring in greater depth the crucial role that human embodiment plays in learning racism—something that is all too often overlooked in the discussion of race and racism.
In the end, the goal of this investigation is not that humans become like robots with respect to racism (since human embodiment makes this impossible), but rather that we might gain greater clarity about the crucial role that embodiment plays in becoming racist. In turn, the hope is that through such clarity we may begin to employ our embodiment towards constructive counter-racism projects and ends.\footnote{As we have pointed out above, our conclusions concerning the inability of robots to learn racism should serve, at minimum, as a good impetus to push us on in pursuing this task.}\footnote{Discussions of racism and robots mostly focus on either human prejudice against machines (see, e.g., Dator) or the possibility of robots being racist against the human race in general.}

\footnote{Or, as I will argue below, a more accurate way to describe learning with respect to racism would be in terms of the acquisition of racism. Learning tends to sound too cognitive, whereas acquisition can incorporate less cognitive or non-cognitive aspects such as habit and skill development. For the purposes of this essay, I blend together the phrases “becoming racist” and “learning racism.”}

Notes

1 Discussions of racism and robots mostly focus on either human prejudice against machines (see, e.g., Dator) or the possibility of robots being racist against the human race in general.

2 Or, as I will argue below, a more accurate way to describe learning with respect to racism would be in terms of the acquisition of racism. Learning tends to sound too cognitive, whereas acquisition can incorporate less cognitive or non-cognitive aspects such as habit and skill development. For the purposes of this essay, I blend together the phrases “becoming racist” and “learning racism.”

3 In section II of this paper I describe in greater depth what I identify as Cartesian ontology. Simply put, this is the substance ontology put forward by Descartes, most characteristically in his *Meditations*, by which he draws a sharp division between the mind and body as two utterly distinct and incommensurable substances while elevating the former over the latter. (On this point, see, for example, Watson.) The suggestion I make in this paper with respect to Cartesian ontology is that such ontology cannot account for the possibility of learning racism. As I argue, not only does this mean that robots, insofar as their ontology is fundamentally Cartesian in nature, are unable to become racist, but any view of racism that likewise depends upon Cartesian ontology (e.g., the view which construes racism simply as a cognitive problem of ignorance or false beliefs that extends from the subject’s cognitive and asymmetrical relation to the world) neither understands what is most characteristic of racism nor is it adequately equipped to confront such racism.

4 Examples of this include F. Bernier’s 1684 *A New Division of the Earth* or S. G. Morton’s 1839 *Crania Americana*. Historically, this position has been associated with attempts to “prove” the otherness and inferiority of non-whites in order to uphold structures of racial inequality. Both of these views are discussed in Taylor 39-42.
Cited in Haslanger 57.

For an interesting account of how racist ideologies as common sense take root in human beings, especially through the family structure, see Lawrence 699-736.

Dreyfus also cites this example in “What Computers Still Can’t Do” 215.

Dreyfus makes the strong claim that it is impossible for robots to cope successfully with common-sense knowledge and he bases his argument for this on the frame problem. A frame is a parameter that one places around a set as a mechanism to help determine the relevancy of facts. For Dreyfus, however, frames do not ease the difficulty facing robots. He notes: “But a system of frames isn’t in a situation, so in order to identify the possibly relevant facts in the current situation one would need a frame for recognizing that situation, etc. It thus seemed to me obvious that any AI program using frames was going to be caught in a regress of frames for recognizing relevant frames for recognizing relevant facts, and that, therefore, the commonsense knowledge storage and retrieval problem wasn’t just a problem; it was a sign that something was seriously wrong with the whole approach” (“Why Heideggerian AI Failed” 248). As further evidence of this failure, we may point to the problems facing mechanized translation programs. The most successful of recent translation programs is Google Translate. What is interesting about this program, however, is that it is not strictly artificial, but rather the translations it makes are based on translation patterns within millions of documents that have already been translated by human beings. In this fashion, Google Translate allows human beings to manage the common-sense work and from this it merely picks up statistically frequent patterns that could become the basis for an appropriate translation.

Or as James R. Mensch puts it, “to copy [human] intelligence—that is, to produce an artificial version of it—the flesh that forms its basis must also be reproduced” (24).

Unless otherwise noted, all page references in relation to Merleau-Ponty’s thought are references to his Phenomenology of Perception.

Some readers have criticized the notion of intentional arc as still too tied to a Cartesian framework. Merleau-Ponty admits as much himself, which is part of the reason that in his later work he describes the relation of the self and the world as a paradoxical intertwining he names “the flesh”. On this point, see Merleau-Ponty, The Visible and the Invisible, especially chapter 4, entitled “The Intertwining—The Chiasm” 130-155. Although The Visible and the Invisible contains significant insight into the issue of ontology, for the purposes of our investigation the Phenomenology of Perception will be more than adequate as a source text.

On this point, see also part 2, chapter 4, “Irish Americans and Race,” in Meagher. I would like to thank Michael Monahan for bringing this to my attention.
For a parallel example from feminist writings concerning bodily comportment and learning of gender norms, see the excellent essay “Throwing Like a Girl: A Phenomenology of Feminine Body Comportment, Motility, and Spatiality” in Young.

We may identify this understanding of racism as the *anaesthetic view* insofar as it downplays or brackets out entirely the role of the body, and thereby the physical senses, in its approach to racism. Because the body is not central in this view, what is emphasized is transformation of mental states through education and theoretical exercise. Racism is seen as ignorance and as such it should thereby be addressed primarily if not solely through provision of correct information. Unfortunately, this approach is certainly not uncommon and extends into many types of psychological anti-racism that include forms of racial awareness training and positive racial imaging, and it is also the model of much of academic philosophy and its overemphasis on conceptual analysis. On this issue, see Bonnet 97-103. For a good example of the premium placed on conceptual analysis by some philosophers of race, see Glasgow.

In an earlier version of this paper, I included a fourth section entitled “Practical Implications: Combating Racism through Embodied Touch” in which I drew together the insights from the discussion of robots and humans in order to make several tentative suggestions for how constructive counter-racism might play out in a concrete fashion. Due to space limitations, I have excluded this section from the present version of this essay. We may, nevertheless, summarize here a few of the section’s major points.

Given the conclusions of our comparison between robots and humans, it is crucial that counter-racism approaches are fundamentally *aesthetic* in the sense that they are based upon embodied touch. But what does this touch look like? How are we to begin with such a project of touch? In most racially charged contexts it may be necessary for touch as a counter-racist approach to begin slowly and progress in an incremental manner, and for this reason we can divide touch into two different degrees with respect to the level of intimacy associated with each.

The first degree is what we may call touch from a distance, and what we may identify with sound. As Anne Fernald has insightfully pointed out, sound plays a crucial role in physically connecting individuals who are not in direct contact through sound waves and the workings of the human ear (Fernald). This is seen especially clearly in how mothers and fathers are able to soothe their children from a distance with their voices by communicating a physiologically gentle form of touch through certain vocal tones. As a starting level in counter-racism projects, sound as a form of touch may have a powerful effect. This will certainly include speaking proper words in proper ways and employing tones that represent warmth and kindness rather than hostility; but for those of the privileged group, this will also mean a willingness to be touched by others through the humble and vulnerable activity of listening. In opening oneself to the physical touch of sound from a person of an oppressed group, even if these sounds are perceived initially as harsh, is to take a crucial step in forming an embodied physical relationship with the other.

Although touch through sound offers a way to begin counter-racism projects, it will only prove effective if it is also paired with the second and deeper degree of touch, namely the robust tactile experience of one body touching another body. This form of touch is much more intimate than the other, and because of this it also carries the potential of being much more problematic. Perhaps the greatest problem with this approach is that the touch of those from a privileged
group may not be welcomed, in which case such touch would serve as yet another form of racial violation or what Sullivan identifies as “ontological expansiveness” (Sullivan 10). How then are we able to hold onto the importance of embodied touch without falling into the problem of physical violation? Although this will likely leave the reader dissatisfied, the answer to this problem will have to revolve around creating a context of physical openness in which reciprocal touch is possible. This may be as simple as placing oneself physically within shared space in a humble manner that intentionally works against ontological expansiveness. Or it may be as complicated as rejecting isolationism by tearing down fences, walls, gates, borders and intentionally blurring the lines between both overtly and covertly racialized spaces. However, most importantly, this openness will have to take the form of an active waiting. The paradigm example I would like to give of this openness and active waiting is the image of the handshake. Not only is the handshake a declaration of peace and disarmament (Wagner-Pacifici 93), but more importantly it serves as an open invitation to reciprocity while creating the possibility for genuine communion. A true handshake can never be forced, but rather by opening one’s hand to the other, the other is empowered to complete reciprocity or to decline it. In this sense, whereas the handshake for the initiator is an act of active passivity, the handshake for the recipient is an act of passive activity. For any counter-racist project to face the deeply rooted problems of racism successfully, the embodied act of the handshake offers a crucial method for doing so. And even though this act alone is insufficient for the massive task of fighting racism, it promises to provide something of a basic foundation on which to counteract the uniquely embodied manner by which humans learn racism and thereby become racist.

16 I am indebted to Michael Monahan for his help in developing these ideas. In addition, I would like to thank two anonymous reviewers for their careful and helpful comments and criticisms. Finally, I am deeply grateful to Helena Nowachek for her invaluable insights on my own racial habits and how these affect, in deep ways, both my American and Swedish personae.

Works Cited


Ng, Roxana. “‘A Woman out of Control’: Deconstructing Sexism and Racism in the University.” *Canadian Journal of Education* 18 (1993): 189-205.


