Computing Machinery and Sexual Difference: The Sexed Presuppositions Underlying the Turing Test

(to appear in *Feminist Philosophy of Mind*, edited by Keya Maitra and Jennifer McWeeny, Oxford University Press; this is the final submitted version)

Amy Kind

In his 1950 paper "Computing Machinery and Intelligence," Alan Turing proposed that we can determine whether a machine thinks by considering whether it can succeed at a simple imitation game. A neutral questioner communicates with two different systems, one a machine and one a human being, without knowing which is which. If after some reasonable amount of time the machine is able to fool the questioner into identifying it as the human, the machine wins the game, and we should conclude that it thinks. This imitation game, now known as the Turing test, has been much discussed by philosophers of mind, and for more than half a century there has been considerable debate about whether it is an adequate test for thinking. But what has not been much discussed are the sexed presuppositions underlying the test. Too often forgotten in the philosophical discussion is the fact that Turing's imitation game is modeled on an imitation game in which a neutral questioner communicates with two different humans, one a man and one a woman, without knowing which is which. In this original imitation game – what I'll call the "man/woman imitation game" – the man wins if he is able to fool the questioner into identifying him as the woman.

Thus arises the question motivating this chapter: How has philosophical engagement with the issue of computer intelligence been influenced by the comparison to sexual differentiation – what I will call "the sex analogy" – on which it is based? In what follows, I begin with two competing interpretations of the Turing test that we find in the literature, one that ignores the man/woman imitation game altogether and another that does not. As I will suggest, however, even on the interpretation that acknowledges Turing's reliance on the man/woman imitation game, the significance of the sex analogy has not been adequately explored. In the second half of this chapter, I thus turn to such an exploration. As we will see, the fact that the Turing test was modeled on a man/woman imitation game seems to have led us astray in various ways in our attempt to conduct an effective investigation and assessment of computer intelligence.

The Standard Interpretation

The original imitation game involves a man and a woman serving as the two contestants plus a neutral questioner who may be either a man or a woman. Turing uses the labels "A" for the man, "B" for the woman, and "C" for the neutral questioner. Each participant is in a separate room and they communicate with one another electronically. C, who does not know which participant is in which room, aims to make this determination. While A's object is to try to fool C into identifying him as the woman, B's object is to try to help C. In describing the kind of exchange that might take place as the questioner attempts to determine which participant is which, Turing supposes that C might ask about the length of the participants' hair. Were C to ask this of A – the participant who, though actually a man, is trying to convince the questioner that he is a woman – Turing reports that A might respond with something like the following: "My hair is shingled and the longest strands are about nine inches long" (Turing 1950, 434; we will return to this specific example later on in the section titled "The Exegetical Question.")

Only once the man/woman imitation game is fully described does Turing move on to introduce his now famous test. The test emerges directly from the original man/woman imitation game with just one simple change: the machine replaces participant A. Interestingly, however, though Turing himself relies heavily on the man/woman imitation game in developing his test, this reliance has been largely ignored in the secondary literature. Looking at various prominent overviews of the Turing test, for example, one finds that they explain the test without any reference whatsoever to the fact that it arises within a context involving a focus on sex (or sex and gender).[1]

Consider, for example, the entry on the Turing test in the *Stanford Encyclopedia of Philosophy*, perhaps the most highly regarded philosophical reference work available today.[2] The fact that the human/machine imitation game is modeled on a man/woman imitation game receives no mention at all when the game is described:

> Turing (1950) describes the following kind of game. Suppose that we have a person, a machine, and an interrogator. The interrogator is in a room separated from the other person and the machine. The object of the game is for the interrogator to determine which of the other two is the person and which is the machine. (Oppy and Dowe 2016, 5)

We see a similar omission in discussion of the Turing test in the relevant entries of the *Routledge Encyclopedia of Philosophy* (Moor 1998), Blackwell's *A Companion to Philosophy of Mind* (Guttenplan 1994), the *Oxford Companion to Consciousness* (French 2009), and many comparable reference works. Even when it is acknowledged that Turing's own description of the game proceeds slightly differently from the description being provided, or in the rare cases when it is acknowledged that Turing's man/machine imitation game is based on a man/woman imitation game, these facts are dismissed as unimportant or irrelevant to the key matter at hand. As French puts it, though Turing provides a description that is "slightly more complicated" than the one that he himself provides, "there is essentially universal agreement that the additional complexity of the original version adds nothing of substance to its slightly simplified reformulation that we refer to today as the Turing test" (French 2009, 642; see also Saygin et al. 2000, 467).

French's claim of nearly universal agreement is perhaps something of an overstatement, as we will see in just a moment. But it is certainly true that such agreement is widespread. Indeed, even Turing's biographer dismisses the original formulation of the game in terms of a man/woman imitation game as a "red herring" (Hodges 1983, 415). We thus arrive at what is often called *the standard interpretation* of the Turing test, an interpretation that treats it as an imitation game in which a computer aims to imitate a human being. According to the standard interpretation, the question at issue for the neutral interrogator is one of *species-differentiation*: Which of the two beings with whom I am communicating is a member of the human species and which is not?[3]

A Minority Interpretation

Against the standard interpretation, a small cadre of philosophers have offered a different interpretation of the Turing test, one that takes the original man/woman imitation game to have considerably more significance.[4] On this minority interpretation, Turing aims not to offer a species-differentiation test but a sex-differentiation test. Just as the man in the original imitation game aims to convince the neutral questioner that he is a woman, so too the computer in the revised imitation game aims to convince the neutral questioner that it is a woman. This interpretation gains at least partial support from Turing's initial description of the test. Consider the key passage from Turing:

> We now ask the question, "What will happen when a machine takes the part of A [the man] in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?" (Turing 1950, 434)

In shifting from the man/woman imitation game to the imitation game involving the machine, Turing notes explicitly that the man is to be replaced with the machine. Since the man's aim in the original game is to fool the interrogator into identifying him as the woman, and since Turing does not say anything about the aim of the game having changed, philosophers who offer this minority interpretation claim that the Turing test is not intended to be a species test at all.

As Saul Traiger argues, referring specifically to the passage from Turing just quoted:

> The interpretation of this paragraph is crucial. We are to substitute a machine, specifically a digital computer, for A in the game. But substituting one player for another in a game presupposes the identity of the game. So nothing about the game has changed. We merely have a different individual, one who happens to be a digital computer, taking on the role of A. The machine, as A, is to cause C to believe that A is a woman. (Traiger 2000, 565)

Likewise, as Judith Genova puts it:

> Obviously, those who eliminate gender from the game assume that when Turing says, "Will the interrogator decide wrongly as often as when the game is played like this", he means by 'this' something like, between a computer and a person, with the object now being to discover which is machine and which is human. However, there is no prima facie reason to change the basic question of the game. 'Like this,' can just as well mean, with the machine imitating a man with the object remaining to distinguish between genders. (Genova 1994, 314)

Thus, on this minority interpretation, Turing did not intend for the interrogator to try to figure out which being is the human but rather to try to figure out which being is the woman.[5] Here the test for intelligence is an indirect one. The machine can be said to pass the test – and thus to count as intelligent – if it manages to fool the interrogator into identifying it as the woman as often as the man had managed to fool the interrogator into identifying him as the woman. The thought seems to be something like this: If a computer is as good at this kind of sophisticated imitation task as a man is, then we can infer intelligence on the part of the computer.

What's the upshot of interpreting Turing in this fashion? Several of the philosophers who offer this minority interpretation take the primary significance to lie in the difficulty of the task facing the machine. That's not to say that they take imitating a woman to be easier than imitating a human of unspecified sex. Rather, the difficulty lies in the fact that in the species-differentiation test the interrogator knows that one of the participants is the machine whereas in the sex-differentiation test, this thought does not even enter the interrogator's mind. (See, for example, Hayes and Ford 1995, 972; Traiger 2000, 565; Lenat 2008, 274.)

Unlike proponents of the standard interpretation, then, proponents of the minority interpretation argue that the man/woman imitation game cannot be ignored when understanding the Turing test. In doing so, however, such proponents deny that we should assign any particular importance to sex. While this may initially seem surprising, it actually coheres nicely with their explanation of the significance of their interpretation. If the important difference between the minority interpretation's view of the Turing test and the standard interpretation's view of the Turing test is that only the former keeps the interrogator ignorant of the fact that one of the contestants in the game is a machine, then there needn't be any special reason that we model the Turing test on a game involving participants of two different sexes as opposed to different nationalities or political alignments. Granted, we may not be able to sub in just any pair of contrasting identities here. Traiger suggests that it be important that there be some sort of cultural alignment among participants (Traiger 2000, 570), while Sterrett suggests that the contrasting identities must be such that the imitator is required to reflect critically on what kinds of responses to give (Sterrett 2000, 550). But the fact remains that, even on the minority interpretation, the focus on sex itself is not seen as an especially important one.

Here Genova's work constitutes a notable exception. As is by now well known, Turing struggled with his sexual identity throughout his lifetime, and this struggle is often thought to have played a direct role in his suicide.[6] Drawing on these biographical details, Genova argues that Turing's work "confirms the belief in a close interaction between the personal and the intellectual affirmed by many feminist and cultural critics today" (Genova 1994, 324). In particular, Turing's use of the man/woman imitation game enabled him to confuse men, women, and machines, and thereby demonstrate "that no boundaries were sacred or unbreachable. All rules, all categories, all boundaries were made to be transgressed" (317). Indeed, Genova sees Turing's inclusion of the machine in the man/woman imitation game as enabling him to deconstruct this binary and, moreover, that "the evolution of thinking machines might provide a solution to the strangling binarism" that comes along with the differentiation between the sexes (315-16).

In this way, Genova sees Turing's initial use of a man/woman imitation game as both deliberate and essential, and she thereby takes a position that sets her apart not only from both the proponents of the standard interpretation but also from other proponents of the minority interpretation. Yet this also reveals something interesting. In bringing sex to the forefront of the discussion, Genova's reflections suggest a way in which our thinking about sex might be influenced by the analogy with computer intelligence. Importantly, however, this still tells us nothing about the ways in which our thinking about computer intelligence might be influenced by the analogy with sex. Thus, even here, in a paper rare among the philosophical literature on the Turing test for its assignment of importance to Turing's invocation of the sexes *qua* the sexes, the question that motivates the discussion of this chapter – how the sex analogy has influenced our thinking about computer intelligence – remains unaddressed.

The Exegetical Question

Now that we have distinguished the minority interpretation from the standard interpretation, the question naturally arises as to which interpretation is the correct one. Though I will not here attempt to settle this exegetical question, it's worth noting some important textual evidence in support of the standard interpretation over the minority interpretation. In particular, it's difficult to make sense of some of the discussion of "Computing Machinery and Intelligence" if Turing did not have a species-differentiation test in mind. Consider, for example, this passage concerning the appropriate strategy for the machine to adopt in attempting to win the game:

> It might be urged that when playing the 'imitation game' the best strategy for the machine may possibly be something other than imitation of the behaviour of a man. This may be, but I think it is unlikely that there is any great effect of this kind. In any case there is no intention to investigate here the theory of the game, and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man. (Turing 1950, 435)

If Turing intended a sex-differentiation test, then the machine (who is taking the place of A, the man, in the original man/woman imitation game) would be aiming to convince the neutral questioner that it is a *woman*. Thus, one would expect to see Turing here talking about the *behavior of a woman* and *answers that would naturally be given by a woman*.

While passages like this are by no means decisive, they are at least strongly suggestive that Turing had a species-differentiation test in mind. But, as noted above, I won't here attempt to settle the exegetical question. Ultimately the question of which interpretation is the correct one is something of a red herring for the issue I here pursue, that is, the issue of how the sex analogy has influenced the subsequent discussion of computer intelligence. Whatever Turing himself intended, the vast majority of philosophical engagement with the Turing test has proceeded against the backdrop of the standard interpretation. Our exploration of the sex analogy must thus operate within that framework, that is, within the framework of the Turing test understood as a species test. This gives us the following formulation of the question in which I am interested: How has our understanding of that species-differentiation test been influenced by the fact that it was modeled on a sex-differentiation test?

As I will argue in what follows, thinking of a questioner's determination of intelligence on the model of a questioner's determination of sex has had several unfortunate consequences for philosophical thinking about computer intelligence. I'll focus on three such consequences in particular:

- The sex analogy has encouraged us to think of intelligence as a certain kind of all-or-nothing, fixed property;
- The sex analogy has encouraged us to think about superficial markers of intelligence; and
- The sex analogy has encouraged us to think of the machine as pretending to be something that it is not.

In the next three sections, I take up each of these consequences in turn.

Intelligence as All-or-Nothing

In thinking about all three of the consequences that I will be exploring, it's important to remember that Turing was operating with a mid-twentieth century conception of one's sex. This point is especially critical in thinking about the first of our three consequences. According to the conception operative in the middle of the twentieth century, people fall into exactly one of two categories when it comes to sex: man or woman. Moreover, which category someone falls into was not considered to be open to change. Though neither of these assumptions currently holds much sway, at the time Turing was writing sex was thought to be both binary and fixed.[7]

In taking the man/woman imitation game to be an appropriate model for a test for intelligence, Turing thereby seems to be thinking of intelligence in an analogous way as both binary and fixed. The very set-up of the test seems to presuppose that there are exactly two categories a machine could fall into. Either the machine is intelligent or it is not intelligent. It also seems to presuppose that which category a machine falls into is not open to change.

In fact, it takes only a moment's reflection to see how problematic these two presuppositions are. In other contexts where we talk about intelligence, we are much more likely to recognize an *intelligence continuum*. Consider discussions of animal intelligence, for example – or more specifically, consider Koko, an eastern lowland gorilla who has been taught American Sign Language as part of her participation in The Gorilla Language Project. According to her handlers, Koko has a working vocabulary of over 1000 signs and is said to understand approximately 2000 spoken English words. On IQ tests, where a score of 100 is considered normal for humans, Koko has scored between 70-95. In discussions of Koko, her mental capacities are often compared to those of a young child (Patterson and Gordon 1993). The discussion is not framed in terms of *whether* she is intelligent, all or nothing, but rather *how much* intelligence we should say that she has. We see a similar dialectic in discussions of intelligence in other animal species. To give just one other example: Alex, an African Grey parrot studied extensively by Irene Pepperberg, is often described as having the intelligence of a five-year old (Pepperberg 2009).

When we step away from a computer-related context to other contexts in which intelligence is discussed, we also see an important sensitivity to the idea that where one falls on the intelligence continuum can change. Animals like Koko and Alex may, with guidance from trainers and teachers, develop new skills and capacities that incline us to count them as more intelligent than they were before. This point is also an obvious one in discussions of human intelligence. Intelligence develops as one grows from infant to child to adult – though also, in some unfortunate cases, accident or illness can lead to intelligence loss. In none of these noncomputer-related contexts do we see an assumption that intelligence is fixed. Why, then, do we make such an assumption in the case of computer intelligence? Why do we see such an assumption built into the notion of the Turing test?[8] As I am here suggesting, this problematic assumption traces directly to the sex analogy.

Having fleshed out the problematic nature of these two presuppositions, however, we are now led to what may seem to be a counterintuitive result. It turns out that it's precisely because the Turing test has been taken to be a species-differentiation test (as assumed by proponents of the standard interpretation), rather than a sex-differentiation test (as assumed by proponents of the minority interpretation) that considerations of sex end up exerting an important influence in the discussion of computer intelligence. To see this, it will be helpful first to recall that proponents of the minority interpretation claim that Turing intended the game to stay the same when the machine substituted in for player A, that is, that we should see the Turing test as a sex-differentiation test. In this way, the neutral questioner is ignorant of the fact that one of the contestants in the game is a machine. As we saw, however, proponents of this interpretation also point out that Turing need not have started with a man/woman imitation game in order to achieve the requisite ignorance on the part of the neutral questioner. A nationality imitation game, or a political alignment imitation game, would have worked just as well. In the context of the standard interpretation, however, the differences between modeling the species-differentiation test on a sex-differentiation test and modeling it on a test aimed at differentiating nationality or political alignment test turn out to be critical ones.

Let's focus on nationality for a moment, and let's suppose that the questioner in the nationality imitation game has to figure out which participant is the American. Here we could have many different kinds of contestants in the role of A: a Canadian, a Mexican, an Australian, a Korean, and so on. When the questioner identifies one of the contestants as being the American, this identification thus carries no further implication about what nationality the other contestant is. Nationality is not binary. The same point would apply to a political alignment game, since political alignment also fails to be binary.

Because these characteristics are non-binary, we also see more clearly a network of overlapping similarities and differences among the various choices. The Canadian and Australian might share a native language with the American, the Canadian and the Mexican share geographical proximity with the American, and so on. On some questions, then, it will likely be easier for the Australian to imitate the American than for the Mexican, whereas on other questions it will likely be easier for the Mexican to imitate the American than for the Australian. Matters are similar with respect to political alignment. With respect to some beliefs a Republican will resemble the Libertarian more than the Democrat, but with respect to other beliefs they will resemble the Tea Partier more than the Libertarian. So again, on some questions it will likely be easier for the Tea Partier to imitate the Republican whereas on other questions it will likely be easier for the Libertarian to do so.

In fact, then, the two-contestant imitation game format seems somewhat unnatural when thinking about nationality or political alignment. If we were starting from scratch in an effort to design a game whose aim was to unmask the American, say, it seems we'd be much more likely to adopt a different format. Perhaps we might use something more like the *Dating Game* set-up, where there are three contestants behind the curtain – one American and two not. Or we might even begin with a much larger group. By careful interaction and well-designed queries, the neutral questioner might eliminate possible contestants one by one until the field is narrowed to just a single contestant left standing – not quite the original *Dating Gam*e, perhaps, but instead more like an anonymized version of its millennial counterpart, *The Bachelorette*.

Even in these kinds of set-ups, there is still only one winner. Only one panelist gets the date, and only one contestant gets the rose. But suppose we were to develop a Turing test based on one of these other kinds of models. I don't know exactly what such a test would look like, but we might naturally think that there would be an important sense in which a computer who placed in the middle of the pack – who came in second among three panelists, or who lasted into a late round in the one-by-one elimination of a larger group – could still be said to be victorious.

The point here should in some sense be obvious: One doesn't have to come in first in an intelligence contest to count as intelligent. But when we're thinking of intelligence on the model of sex, this point gets lost. When the model at work is one of sex, and when sex is taken to be all-or-nothing, intelligence too looks like an all-or-nothing property. Indeed, intelligence looks like the kind of property that species-membership is. Either you are a human or you're not, and likewise, either you're intelligent or you are not. In neither case is there any middle ground. In this way, the test seems to encourage a way of thinking about intelligence that unnecessarily prejudices us against the computer participant.

Unsurprisingly, perhaps, many of the influential criticisms of the Turing test in the philosophical literature can be seen as pointing to something like this prejudice. To take just one example, consider Robert French's argument that "the Test provides a guarantee not of intelligence but of culturally-oriented human intelligence" (French 1990, 54). In an effort to establish this point, French offers the analogy of what he calls "the seagull test" for flight: a machine is said to be capable of flight if a neutral observer is unable to differentiate the machine from a real seagull by way of a three-dimensional radar screen. While it seems true that any system that passes the seagull test is flight-capable – that is, while it seems true that passing the seagull test is sufficient for flight-capability – it also seems true that many flight-capable systems will be unable to pass the test, as the test sets the bar too high. The point here seems clear: Just as flight-capability should not be linked to being indistinguishable from a seagull, so too should intelligence not be linked to being indistinguishable from a human. To assume otherwise is to be humancentric. But while this kind of humancentric prejudice has been previously noticed, what hasn't yet been noticed is how such prejudice can be seen as a natural consequence of the sex analogy with which Turing began.[9]

There's a certain irony here, since one of Turing's main aims in proposing the test was precisely to sidestep a certain kind of prejudice that would arise from tackling the question of computer intelligence head on. Given a mid-century definition of "thinking," Turing worried a computer might be conceptually excluded from counting as thinking. The concept "thinking" might simply not have been a concept that could be applicable to a computer. The Turing test was a way to circumvent this problem, or at least, a way to postpone direct consideration of the question, "Can a computer think?" until such time as the definition of thinking had become sufficiently expansive to allow for a genuine assessment.[10] In modeling his test on a man/woman imitation game, however, Turing showed himself to be victim of some of the same prejudices that he was trying to avoid.

Superficial Markers of Intelligence

The previous section helped us to see how the sex analogy misleadingly encourages us to think of intelligence as all-or-nothing. In this section, we will look at a second way in which the sex analogy is misleading. But while the first problem arises primarily from a way in which the sex analogy encourages a problematic characterization of intelligence itself, this problem arises from a way in which the sex analogy encourages a problematic way of detecting intelligence. In particular, the sex analogy seems to encourage a superficial line of questioning by the neutral questioner, and it thus correspondingly encourages us to focus on superficial rather than deep markers of intelligence.

Recall the kinds of question that Turing thinks the neutral questioner might ask in the original man/woman imitation game, namely, questions about hair length. Even from a 1950s perspective one would expect that hair length and other matters of appearance would be viewed as a superficial marker of sex and not a central aspect of it. In fact, the contestants are separated from the neutral questioner precisely so matters of appearance don't influence the questioner's judgment.

Of course, the questioner might choose to probe in a different way. By forcing the contestants to answer questions about moral dilemmas or about matters of social justice, a questioner might be able to uncover subtle patterns of sexed thinking lurking beneath the surface. Insofar as an awareness of such patterns requires a nuanced understanding that would be difficult for an unsophisticated participant to fake, we might think that these lines of questioning would ultimately prove more fruitful than lines of questioning about more superficial matters. But these lines of questioning are not among the most obvious ones that come to mind. Rather, in an imitation game involving sex, we're likely to be inclined toward questions about appearance, occupation choice, favorite pastimes and so on – an inclination that is promoted by Turing's own choice of sample question.

Ultimately, though Turing could perhaps have chosen to highlight more illuminating examples of sex-related questions, to some extent the problem here may be largely unavoidable, that is, the focus on superficial markers may well be endemic to the very concept of a sexed imitation game. Much contemporary work on sex and gender suggests that many of the apparent differences between the sexes are not biologically based but rather are the products of various kinds of socialization. As such, these differences are at least in principle changeable and thus might not be thought to run very deep. But insofar as there are no deep facts about sex that could manifest in a conversation with a neutral questioner, there are no deep facts for which to test. In this respect, sex seems quite unlike intelligence. Whether a being has intelligence does seem like a deep fact about the being. Were we not thinking about testing for intelligence on the model of testing for sex, we might naturally focus our attention more effectively on rooting out that deep fact. With sex as the model, however, our attention is naturally diverted to more superficial markers. Worries, for example, are explicitly raised in Turing's discussion about how fast or how flawlessly the computer might answer questions about calculation. Too quick or too flawlessly, and the computer might be inadvertently unmasked. Here we bump up against a certain irony: it's the computer's *superiority* with respect to these superficial markers of intelligence that might keep it from passing the test.

This irony brings us back to an issue that we encountered in the previous section: Turing's species test seems to constrain intelligence to a human mold.[11] And here again, we can see how the sex analogy helps to impose this constraint. To convince a questioner that they are a woman, participants in the man/woman imitation game will likely fall back on traditional conceptions of femininity and womanhood. This will be true not only for contestant A, who is not the woman, but also for contestant B, who is. Suppose contestant B is 5 foot 10 inches tall or has very short hair or that she has some other physical characteristics that are not traditionally viewed as feminine. Or suppose that she is a race car driver or an actuary or a member of some other profession dominated by men. Speaking truthfully about any of these characteristics will likely make it harder for B to win the game. In this way, the man/woman imitation game constrains sex to a very traditional or stereotypical – indeed, superficial – mold. It's thus no surprise that a species test modeled on it would likewise constrain intelligence to a very traditional or stereotypical – indeed, superficial – mold. The ability to answer questions quickly and flawlessly matters no more to intelligence than profession and hair length matter to sex. But, in both cases, it's these kinds of superficial characteristics that end up looming large.

At this point, however, an objection to my line of reasoning might naturally arise. In describing the Turing test, Turing himself gives three different examples of questions that might be put to the participants. The first asks the participant to produce a sonnet, the second asks the participant to solve an addition problem, and the last asks the participant to solve a chess problem. Since facility with poetry, math, and/or chess doesn't seem in any way a superficial marker of intelligence, the point pushed in this section might well seem unfounded.

In response to this potential objection, I think it's useful to return to a point that I made in the previous section, namely, that the notion of intelligence being employed in discussions of computer intelligence seems quite different from the notion being employed in other contexts, for example, in discussions of animal intelligence. The fact that Koko the gorilla and Alex the parrot cannot compose poetry or play chess does not count against their having intelligence. Passing the Turing test seems to require not only being able to respond like a human, but being able to respond like a human over an extremely wide range of subjects. Though having such a wide-ranging capacity for response might well be a good indicator of intelligence, it is by no means essential to intelligence, and a focus on it moves us away from the deep facts about intelligence, whatever those may be.

The Pretense of Intelligence

In addition to the two problematic influences already discussed, in this final section I'll briefly consider an additional problem that arises from the kind of contrast that the sex analogy sets up. In the man/woman imitation game, the man is pretending to be something that he is not. The whole point of the game is for him to try to fool the questioner into identifying him as something that he's not. If he wins the game, we certainly aren't entitled to conclude that he is a woman but rather that he's a good pretender – or at least that he's a good pretender when it comes to his sex. This encourages us to think about the computer's performance the same way: Just as the man is pretending to be something that he is not, the model of the man/woman imitation game thus encourages us to think of the computer participating in the Turing test as pretending to be something that it is not. If the computer wins the game, why would we conclude that it is intelligent? Wouldn't the appropriate conclusion simply be that it is a very good pretender, analogous to the appropriate conclusion in the man/woman imitation game?[12]

In fact, the issue of pretense has played a significant role in criticisms of the Turing test. Consider, for example, the so-called Chinese Room thought experiment proposed by John Searle. Searle offers a case in which a system might pass the Turing test for speaking a language like Mandarin without having any understanding of Mandarin, a fact that each of us is supposed to be able to determine for ourselves by imagining ourselves instantiating a computer program for the production of Mandarin outputs (Searle 1980).

At this point, one might plausibly protest that the problem of pretense arises not from the fact that Turing modeled his test on an imitation game involving sex but that he modeled it on an imitation game at all. In any imitation game, the issue of pretense will be a salient one. Worries about pretense – that the computer was not actually intelligent but was merely pretending to be so – would have just as naturally arisen had Turing chosen to focus on political alignment of nationality. While this point seems undoubtedly correct, I don't think it undermines the point that I mean here to be making. For I think there's good reason to believe that it's the focus on sex that leads us to think about imitation games in the first place.

The philosophical exploration of computer intelligence prompted by Turing's introduction of his test is by now almost seventy years in. And while the discussion has in many ways moved away from the test itself, with numerous commentators vehemently arguing that it has led the conversation astray, until now little attention has been paid to the respect in which the sex analogy underlying the test has played a part in that conversation. As I have aimed to show in this chapter, its part has been not only significant but also in various ways problematic. Of course, it's impossible at this point to unwind the clock and wipe the slate clean. But even if we cannot entirely erase the problematic influences of the sex analogy, we can only begin to address them and work toward a corrective once they have been identified. Though there is thus considerably more work to be done, this chapter should serve as an important first step towards this ongoing project.

References

Fausto-Sterling, Anne. 1993. "The Five Sexes: Why Male and Female Are Not Enough." *The Sciences* (March/April): 20-25.

French, Robert. 2009. "The Turing Test." In *Oxford Companion to Consciousness*, ed. Timothy Bayne, Axel Cleeremans, Patrick Wilken, 641-643. Oxford: Oxford University Press.

French, Robert. 1990. "Subcognition and the Limits of the Turing Test." *Mind* 99: 53-65.

Genova, Judith. 1994. "Turing's Sexual Guessing Game." *Social Epistemology* 8: 313-326.

Guttenplan, Samuel. 1994. "Alan Turing." In *A Companion to the Philosophy of Mind*, ed. Samuel Guttenplan, 594-596. Oxford: Blackwell.

Hayes, Patrick and Ford, Kenneth. 1995. "Turing Test Considered Harmful." *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 972-977.

Hodges, Andrew. 1983. *Alan Turing: The Enigma.* New York: Simon and Schuster, Inc.

Lenat, Douglas B. 2008. "Building a Machine Smart Enough to Pass the Turing Test: Could We, Should We, Will We?" In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, ed. Robert Epstein, Gary Roberts, and Grace Beber, 261-282. Dordrecht: Springer.

Moor, James H. 1988. "Turing, Alan Mathison (1912–54)." In *The Routledge Encyclopedia of Philosophy*, ed. Tim Crane. Taylor and Francis. https://www-rep-routledge-com.ccl.idm.oclc.org/articles/biographical/turing-alan-mathison-1912-54/v-1.

Mikkola, Mari. 2017. "Feminist Perspectives on Sex and Gender." In *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), ed. Edward N. Zalta. <https://plato.stanford.edu/archives/sum2017/entries/feminism-gender/>.

Oppy, Graham and Dowe, David. 2016. "The Turing Test." In *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), ed. Edward N. Zalta. <https://plato.stanford.edu/archives/spr2016/entries/turing-test/>.

Patterson, Francine and Wendy Gordon. 1993. "The Case for the Personhood of Gorillas." In *The Great Ape Project*, ed. Paola Cavalieri and Peter Singer, 58-75. New York: St. Martin's Griffin.

Pepperberg, Irene M. 2009. *Alex and Me*. New York: Harper.

Proudfoot, Diane and Copeland, Jack. 2012. "Artificial Intelligence." In *The Oxford Handbook of Philosophy of Cognitive Science*, ed. Eric Margolis, Richard Samuels, and Stephen P. Stich, 147-182. Oxford: Oxford University Press.

Saygin, Ayse Pinar; Cicekli, Ilyas; and Akman, Varol. 2000. "Turing Test: 50 Years Later." *Minds and Machines* 10: 463-518.

Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3: 417-57.

Sonnad, Nikhil. 2015. "This Free Online Encyclopedia Has Achieved What Wikipedia Can Only Dream Of." Available at https://qz.com/480741/this-free-online-encyclopedia-has-achieved-what-wikipedia-can-only-dream-of/

Sterrett, Susan 2000. "Turing's Two Tests for Intelligence." *Minds and Machines* 10: 541-559.

Stoller, Robert. 1968. *Sex and Gender.* London: Hogarth Press.

Traiger, Saul. 2000. "Making the Right Identification in the Turing Test." *Minds and Machines* 10:561-572.

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* 59: 433-460.

An earlier version of this chapter was presented at a session on "Feminist Philosophy of Mind" at the Pacific Division meeting of the American Philosophical Association in 2017. I am grateful to the audience there for helpful feedback. Thanks also to Keya Maitra and Jennifer McWeeny for their incisive and helpful feedback.

[1]Though I here say "sex and gender," it is important to note that Turing was writing at a time before the sex/gender distinction was in place. (The distinction is often traced to Stoller (1968); see Mikkola (2017) for a useful overview.) In order to avoid anachronism, and also to better capture what Turing had in mind, I tend to talk in terms of "sex" rather than "gender" in what follows.

[2] For a discussion of the prominence of the Stanford Encyclopedia of Philosophy, see Sonnad 2015.

[3] In addition to the writers mentioned in the text above, the standard interpretation is also offered by Jack Copeland and Diane Proudfoot, two philosophers who have (both jointly and individually) written extensively on Turing's work. See, for example, Copeland and Proudfoot (2012). See also Moor (1976, 249).

[4] See Genova (1994), Hayes and Ford (1995), Sterrett (2000), Traiger (2000), and Lenat (2008).

[5] Interestingly, Lenat suggests that we can attribute the dominance of the standard interpretation to "political correctness." See Lenat (2008, 273-4). I do not have the space to engage this suggestion here.

[6] See Hodges (1983) for extensive discussion of Turing's struggles with his sexuality.

[7] Indeed, the view that gender is both fixed and binary was still the prevailing view in the 1990s. In pushing back against this binarism in her famous article "The Five Sexes," Anne Fausto-Sterling notes that "Western culture is deeply committed to the idea that there are only two sexes" (Fasuto-Sterling 1993, 20).

[8] Of course, subsequent to the publication of Turing's paper, there has been considerable work in computer science and artificial intelligence towards enabling machines to learn. The existence of such work, however, does not undermine my point in the text above about the presuppositions built into the Turing test. Whatever the progress on learning computers that has been made, such presuppositions have nonetheless exerted an influence in the philosophical exploration of computer intelligence.

[9] French, for example, argues persuasively that part of the problem with the Turing test is that it "admits of no degrees" in its determination of intelligence and that there is a problem with any kind of test that fails to recognize an intelligence continuum (French 1990, 57). But in making this argument, French does not seem to notice the way in which this failure on the part of the Turing test seems to trace to the sex analogy in which it is based.

[10] See, for example, the following passage: "The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (Turing 1950, 442).

[11] In fact, the constraint is even narrower than this, since there seems a presumption that the human contestant is a Western, neurotypical adult.

[12] In fact, this issue of pretense might well be one reason why the model of the man/woman imitation game often drops out of the discussion. Insofar as it seems plausible that simulating intelligence may be good enough for intelligence, while still remaining implausible that simulating being a woman is good enough for being a woman, we have reason to think that the analogy should be dismissed as irrelevant – as just a way to get the conversation started.