What Do Gendered Machines Mean?

Gender Representation and the Dynamics of Representational Claim-Forming in Intelligent Machines

by

Nicole Bradley

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved July 2021 by the
Graduate Supervisory Committee:

Cheshire Calhoun, Co-Chair
Ben Phillips, Co-Chair
Peter de Marneffe

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

Technology has a representation problem. While, in recent years, much more attention has been given to how developing technologies exacerbate social injustices and the marginalization of historically oppressed groups, discussions surrounding the representation of marginalized voices are still in a somewhat nascent state. In pursuing a future where underrepresented groups are no longer underrepresented (or misrepresented) in technological developments, I use this thesis project to draw attention to how gendered technologies are said to represent women as a class. To frame the sort of representation problem I have in mind here, I explore the dynamics of representing others as being a certain way, how individuals can be justified in their practice of representing others as being a certain way, and how such representations might produce harm. I draw special attention to particularly controversial technologies such as Sophia the Robot and sexbots in order to address issues of accountability and dehumanization. I end with some, perhaps, encouraging notes about how the sort of responsible design practices outlined in my project might open the door for some compelling liberatory developments.

DEDICATION

To my (rad) dad, for raising me to treat the world with care and kindness.

ACKNOWLEDGEMENTS

I'm here, I've accomplished much, and thus I'm seriously in debt. Where to begin in recounting my debts?

I owe my partner Blake and our dog Lyra for the love and warmth that have helped me to survive. They've given all of my actions meaning and, just by existing, made the world all the more enchanting.

I owe my family, Chris and Kyley included, for the openness and respect that has helped me become the individual I am today.

I owe Alex Criddle for being my official friend and unofficial advisor.

I owe the entire graduate philosophy student cohort for *always* being welcoming and willing to lend a hand.

I owe past educators who have encouraged me and broadened my understanding of what exists and is possible. I owe a few of them, in particular, for pushing me to do the things that terrify me.

Lastly, I owe my thesis committee for their support, guidance, and resources.

TABLE OF CONTENTS

Page

CHAPTER

CHAPTER Page

Sexbots, Women, and What It Means to "Represent"

I.I. Introduction

I'm interested in gendered machines and what they seem to say about the gender groups they're purported to represent. Section I of my larger project intends to explore what "representation" in these cases means, what it involves, and how we try to make sense of it. For this discussion, I will make frequent reference to sexbots as an exemplar of the sort of representational (gendered) machine I'm interested in. Section 3 will add further detail to sexbot-specific representation, but for this first section it will suffice to say that sexbots are a frequent target of feminist technology critique; for some they are pornography made physical and in that way they carry with them all of the harms pornography was purported to carry during the anti-porn/sex-positive debates. Common criticisms of sexbots see the root of sexbots' harm as being their representational nature; like women depicted in porn, sexbots "represent" women as a class. Through sexbots' representations of women, they characterize women as being a certain way—sexually compliant, eager to please, sexual objects, and so on. In what follows of Section I I explore the complexities of representation to better understand what a statement like "Sexbots represent women as readily available sexual objects" expresses about sexbot manufacturers, audience reception, and the process of forming representational claims.

The sort of representation I'm most interested in is that which is propositional and characterizes people, specifically traditionally marginalized peoples, as being a certain way to an audience. Such a representation might come in the form of verbal descriptions, media depictions and artistic portrayals, and—increasingly so—technological artifacts that to some extent 'stand in for' and characterize their target groups. There are often serious sociopolitical consequences to

these representations. To characterize a group of people as being a certain way is to make a claim about what they are like; representational acts make claims about their objects. These claims might guide audiences to perform acts in response to what the characterized object is said to be like; "Group G is dangerous" implies that the audience should use caution around that group, and similarly "Group G wants equal pay" will prompt different actions depending on the audiences being addressed (think addressing an employer who can adjust pay rates versus addressing coworkers outside of Group G who could support the equal pay initiative). How these characterizations motivate audiences to act, particularly in cases where there seems to be much at stake for the characterized group, creates a pressure to characterize others with care.[1]

The characterizations do not directly and seamlessly transfer from the speaker's mind to the audience's mind(s); the "meaning" of a representation must be thought through and interpreted. Here's a sticky point. We factor in the context in which a representational act takes places, our own personal experiences, our cultural backgrounds, our understanding of social meaning and meaning-making practices, to interpret the representations we are confronted with.[2]

---

[1] Put another way, if we care about how vulnerable populations are systemically disadvantaged/empowered, what opportunities are afforded to them based on common conceptions of them as a group, and what dangers exist for these groups when they are thought of in certain ways, then we should be concerned about how we characterize these groups to others who might use such characterizations in order to decide how they should act toward the characterized groups.

[2] See also Sally Haslanger "Social Meaning and Philosophical Method," *American Philosophical Association 110th Eastern Division Annual Meeting* (2013), http://web.mit.edu/~shaslang/papers/SMPMhdo.pdf. Sally Haslanger describes social meanings as being "constituted by the schemas that interpret resources for us," these schemas "consist[ing] in clusters of culturally shared concepts, beliefs, and other attitudes that enable us to interpret and organize information and coordinate action, thought, and affect" (1-2). Borrowing from Lawrence Lessig's work on social meaning, Haslanger adds that these schemas are anchored to the social context in which meanings are interpreted and produce a range of possible meanings, the boundaries of such a range changing over time as the social contexts themselves evolve. This is similar to what I have in mind when I discuss social meaning. To clarify how social contexts work in this way, we could reflect on how the message "Know your place" takes on new associations and understandings in the social contexts of, first, an employer reprimanding an employee for issuing demands to another department versus, second, a husband reprimanding his unemployed wife for providing input on

And as persons of different cultural backgrounds, subscribing to different sets of beliefs and socially shared meaning-making practices, encounter the same representation, there is no reason to suspect that they will interpret the representation in the exact same way.

Commonly, disagreement about how "meaning" can be derived from a representation rests between, first, the idea that meaning is imposed on the representation by the speaker and, second, that meaning is interpreted and thus fixed by the audience. Critics of sexbots, like anti-porn scholars, have often placed meaning-making in the hands of audiences. For these critics, "representation" is a stable, fixed act whose results vary only in so far as audiences interpret the meaning of that representation in different ways. But what do we give up when we frame our concerns about "representation" in this way? My suspicion here is that we give up more than we mean to—a focus on the audience-fixed interpretation of meaning compels us to ignore the process of forming representational claims, and it strips us of the tools needed to assess a speaker's accountability when they form damaging representational claims. However, while we want to hold speakers accountable, we must also consider that audience interpretation (given its correlation with action as well as an awareness that marginalized groups have much to lose from negative representations) carries its own special weight. Each group, the speaker(s) and the audience, has a hand in determining what a representational act *does*, and so we might think that each group carries a normative obligation to perform their role (speaker or audience) well.

---

the husband's spending habits. In the first case, knowledge of workplace conduct, outlined job roles, and employment hierarchies do much of the work in delineating what the message means for the reprimanded employee. In the second case, the wife might consider the marriage dynamics, gender roles (especially those concerning wifely duties), cultural attitudes about "breadwinners" and what their financial support obliges them to do, and so on. In the second case, there could be widely different beliefs and attitudes factored into the interpretation if the wife were part of a culture or religion that saw all finances and shared equally between partners no matter who was the top earner.

The anti-porn/sex-positive debate created an awareness of some of these issues and the widespread social injustice that could potentially follow from problematic representations. I see gendered technologies like sexbots and gendered virtual assistants as continuing this debate while creating spaces for questions like 'What are the dynamics of representational claim-making?' and 'How might a speaker be justified to characterize a group as being a particular way?' I also see gendered technologies as posing some unique questions: How does sex with a robot represent sex with women? Is there a special, further, unique harm in the way fembots[3] are said to represent women? If there is a unique harm in fembots representing women, one separate from issues of pornographic and problematic artistic representations, what special obligations does a bot designer have when considering how they ought to design bots? In what follows of Section I, I will focus on the dynamics of representational claim-making, the problem of deriving meaning from representations, and normative principles that should guide representing others as being a certain way. Once we get a handle on these questions, we'll be equipped to examine gender representation in machines.

I.II What Are We Doing When We Represent Others as Being a Certain Way?

Michael Saward's "constitutive" view of representation bridges a well-discussed divide between aesthetic and cultural representation and political representation.[4] Saward's focus is on the creative process of forming representation claims, the dynamics of this representation-forming, ways the represented are "constructed" through representations, and the contestability

[3] "Fembots" here is an inclusive term that describes both sexbots and female-modeled robots not to be used for sexual purposes.

[4] Michael Saward, "The Representative Claim," *Contemporary Political Theory*, vol. 5 (2006): pp. 297–318, https://doi.org/10.1057/palgrave.cpt.9300234.

of representation claims. Saward writes with force that "We need to move away from the idea that representation is first and foremost a given, factual product of elections, rather than a precarious and curious sort of claim about a dynamic relationship."[5] Though Saward has particular interest in "political representation," Saward makes it clear that rather than categorizing the different forms of representation (that is, distinguishing between an elected official representing a part of the population or the leader of a human rights movement representing the movement's interests), discussing the fundamental structure of representation claims allows for more involved detailing of the roles involved and what should be expected of those roles. Saward's account promotes attention to the interactivity of the roles involved in representational claim-making.[6] The maker is engaged in the expression; the audience is engaged in making sense of it. But that's not the end of the story. Contestability is a key component in this dynamic, which I intend to flesh out more fully as this paper develops. Clarity on the roles involved, which Saward sets the stage for, will open up space to better understand the disagreement between the anti-sexbot claim and the feminist sexbot claim and why contestability matters for this disagreement.

In their arguments about what sexbots represent, anti-sexbot thinkers often disregard the process of forming claims and disregard the roles of both the representation owner (the speaker, the claim-maker) and the audience's uptake of the representation, framing the representational content of sexbots as being a product of shared social meaning that worms its way into our

---

[5] Ibid., 298.

[6] Though I ultimately adopt Thomas Fossen's triadic account of claim-making dynamics, it might be good to acknowledge that Saward's account of the dynamical structure is this: "A **maker** of representations (**M**) puts forward a **subject** (**S**) which stands for an **object** (**O**) which is related to a **referent** (**R**) and is offered to an **audience** (**A**)." Ibid., 302.

psyches and compels us to act in socially destructive ways. If representation meaning rests in the shared social meaning, then the process of forming claims, the dynamics of claim-forming, the roles involved in claim-forming, are irrelevant. Feminist sexbot thinkers, on the other hand, argue that the social position of the representation owner (hereafter, simply "RO") and the RO's intentions color the meaning of a sexbot's representational content. Moreover, shifting socially shared meanings and values color the audience's uptake of that representational content. The Sawardian attention to the dynamics of representation aren't outright neglected here, but they are in want of better description.

To be clear, both anti-sexbot and feminist sexbot proponents agree that female-modeled sexbots characterize women in some way; sexbots are being presented to an audience as representing women in some way. But the anti-sexbot account neglects the role of the RO (in favor of social systems that produce this representation and its ability to be understood as having so and so meaning), and also neglects the engagement of the "audience," while keeping intact the subject representing an object in such and such ways to a (passive) audience. Anti-sexbot thinkers think the characterization is implicitly socio-historically informed but explicitly and (nearly) universally understood by the audience (the designers' intentions and perhaps accountability are neglected here). Feminist sexbot proponents recognize the engagement of the RO and audience and think that the characterization changes based on cultural ideas and the speaker's position/context. Additionally, in the feminist sexbot account, the meaning of that characterization is subject to change as the audience shares in the mutation of cultural ideas about sexuality. Feminist sexbot proponents, like sex-positive feminists, think that the historical "representations" present in the sexbot industry (as in the porn industry) are problematic, but that

new representations can and should replace them so long as those new representations can speak *for* the represented class's interests in productive and beneficial ways.

But often sexbot designers are not speaking *for*—they are characterizing without considering the interests of the represented. They are, to use Linda Alcoff's distinction, speaking *about* but not speaking *for*.[7] Speaking *for*, as Alcoff puts it, is a specific subset of speaking *about* activities.[8] I will return to Alcoff's distinction shortly, but the central issue of speaking *about* without considering the interests of the represented presents the need for a normative criterion for representational acts. Thomas Fossen's account of *representation as* rivals Saward's account of representational claim-making dynamics in its specifics while emphasizing the necessity of normative principles for representational acts.[9]

Fossen's account of representation *as* consists of a triadic relationship between a "subject,"[10] an "object" (characterized in some way), and an "audience," but this triadic

---

[7] Linda Alcoff, "The Problem of Speaking for Others," *Cultural Critique,* no. 20 (1991-1992): pp. 5-32, https://doi.org/10.2307/1354221.

[8] Ibid., 9-10. Alcoff divorces representation from "speaking for others." Alcoff thinks that "the issue of speaking for others is connected to the issue of representation generally, the former. . . a very specific subset of the latter" but is "skeptical that general accounts of representation are adequate to the complexity and specificity of the problem of speaking for others." (10). I think Alcoff is correct that speaking for others is a specific subset of representation. What I take in Alcoff's account to be the defining difference between general representation and specifically speaking for others is that speaking for others takes up the representation act in order to achieve some political end on behalf of the represented; this may entail assuming that the represented are not positioned to speak for themselves and/or be heard by the audience of the representation. Fossen's representation also has this political agenda. My interest is in the socio-political representation of women. For these reasons, I personally collapse "speaking for others" and "representation as."

[9] Thomas Fossen, "Constructivism and the Logic of Political Representation," *American Political Science* 113, no 3 (2019): 824-837, https://doi.org/10.1017/S0003055419000273.

[10] Where Fossen's "subject" is an agent making the representational claim, they are what I refer to later as the "representation owner." Where the subject is not an agent, this is not the case. More on this follows.

relationship account does not neglect the engagement of the subject and audience.[11] The subject (X) represents the object (Y) as (Z) to some audience. Often, the subject is a moral agent representing another person or group of persons as being something (interested in I, having the value of V, some characteristic C, etc.). The subject (here, a moral agent) is presumed to be justified in some way to make claims about what Y is and what Y's values and interests are. An example of this triadic representation might be the President of the Gamer's Club (X) characterizing members of the Gamer's Club (Y) as feeling disrespected (Z) by the Student Activities and Extracurriculars Committee to whom she is addressing her statement. The subject's right to speak on the object, to represent the object in some way, can of course be called into question, which leads to a guiding normative principle articulated by Hanna Pitkin which Fossen attempts to recover: responsiveness.[12]

When representing others, we must be responsive to them contesting the way we represent them. I intend to address this more fully in a later section entitled "Justification." For now, it is enough to highlight the dynamics of responsiveness as Fossen explains them. Briefly, the represented are the at-risk party here and, as such, obtain a certain entitlement to a justification of the speaker's representational claims (lest their interests be threatened), whereas the speaker is not entitled to similar justifications of the represented's actions.[13] The speaker is obligated to represent the represented in ways that align with the represented's interests—and

---

[11] That triadic relationship accounts often pose a lack of engagement among the "maker" and "audience" in the claim-making process is a concern of Saward's.

[12] Fossen, "Constructivism," 834.

[13] Ibid.

representations can always be contested.[14] Responsiveness is a normative principle, but it also clarifies how the dynamics of representational claim-making should look. The speaker has a special obligation, which the represented can hold them to. So, how does this responsiveness relate to (or differ between) speaking *about* and speaking *for*?

For Alcoff, the distinction between speaking *about* and speaking *for* is that speaking *for* involves representing others in the interest of the represented; when speaking for others, we take up the responsibility of speaking for others when it seems appropriate that we speak in lieu of their own testimony.[15] The dynamics of speaking *for* representation claims are the same as speaking *about*, but there is a tempting move at this point of distinction that can muddy our moral responses to speaking *about* claims. We might think that speaking *for* claims require RO responsiveness in a way speaking *about* claims do not. I think this is the wrong move. Whether an RO forms a misguided and harmful representation of a certain group of individuals with that group's interests in mind or without considering their interests, the RO is still accountable for those claims and should still be responsive to any contestation the harmed group throws their way. Responsiveness must still be a normative principle guiding speaking *about* claims, so speaking *about* claims are still subject to a bi-directional flow of feedback between the RO and the represented. Contestability and moral obligation to make well-informed, non-harmful representation claims remain significant normative principles guiding speaking *about* claims. Given the moral obligation to consider the represented's interests in both speaking *for* and speaking *about* claims, we might wonder if there really is such a distinction between speaking

---

[14] Ibid., 834-835.

[15] More on this follows.

for and speaking *about* after all. Both seem guided by the same normative principles and carry similar consequences for misrepresentation. Diving deeper into this would distract from the central concern here, but for now we can say that speaking *for* and speaking *about* have the same relationship dynamics, similar normative principles, and similar social consequences when done poorly; for these reasons, we might do well to collapse them for our purposes under "*representation as*" as Fossen describes it.

I.III Representation Meaning

Now that we have set up the players, roles, and rules of representation to be analyzed, we can move to discussing the meaning of the representation. Let's start with a case:

> Say a sexbot company wants to design their sexbots to look like real human women and to communicate something about women's proper place in society. The company programs the bots to be submissive, look vulnerable, not speak unless the customer requests it, and offer a large variety of services catered to the customer's interests. The company wants their sexbots to represent women as submissive and servile to their customers.

Much design work will be needed to accomplish this goal, especially if the meaning of the representation does not rest squarely within the intention of the representation owner (RO). Put another way, the company would need to go a few extra steps to ensure the representation they intend is understood by their target audience. If there is a mismatch between the intended meaning of the representation and the audience's interpretation of the meaning of the representation, what does that representation mean?

In "Robots, Rape, and Representation," Sparrow adopts the view that (most) sexbots represent women (per their being modelled on mostly women) and that sex with a sexbot

represents sex with a woman.[16] According to Sparrow, the intention of the RO does not fix the meaning of the representation. Rather, Sparrow says, "Because meaning is social, the representational content of symbols or actions is determined by reference to the understandings of the relevant community."[17] Sparrow's assessment raises some difficult problems centered around the tension between the norm-divergent agent and the moral community, however.

Generally, a relevant moral community might be ignorant about some of the salient details surrounding an agent's actions. Take a more innocuous example of the tension here: in the United States, most native-born citizens have a deeply engrained perspective on punctuality and what tardiness represents. If someone is 10 minutes late to meet you, you might be annoyed and think something along the lines of "This person doesn't seem to respect me or my time; they'd better have a good excuse for being late." Tardiness here is assumed to represent something like disrespect or apathy toward you, the person left waiting. Alternatively, you might think that tardiness here represents the tardy person's unreliability. In either case, what the tardiness represents is presumably determined by the norms of the relevant moral community. Now imagine that the tardy person is from a culture where there are no strict norms about punctuality; the culture they are from is a relaxed one which values patience and being easy-going and tardiness does not at all represent disrespect. Perhaps the tardy person has not adapted to U.S. culture and all of its norms surrounding punctuality. Is it fair to say that their tardiness

---

[16] Robert Sparrow, "Robots, Rape, and Representation," *International Journal of Social Robotics* 9, no. 4 (2017): 471, https://doi.org/10.1007/s12369-017-0413-z. Sparrow notes that there are male models in the sexdoll industry, but that they are a minority. Referencing data on gendered differences on perceptions of sexdolls and sexbots and product consumption already existing in the industry, Sparrow identifies the market for these products as dominated by men. For this reason, Sparrow assumes the subject of this paper to be feminine sexdolls and male consumers.

[17] Ibid., 13.

represents disrespect or unreliability in this case? If we agree with Sparrow that what a symbol or action represents depends on the understanding of the relevant moral community, we would have to say that the tardiness does represent disrespect or unreliability unless we take one of two routes forward: (1) we could say that the tardy person is not a member of the relevant moral community and so their actions cannot be appropriately processed by that moral community or (2) that the relevant moral community in this case is not ideal and that their ignorance is in need of correction.

The ruling of the ignorant moral community can be particularly unjust in cases where the community's ignorance about topic T is constructed or the community generally holds prejudiced views about topic T. Where ignorance is constructed and upheld, like in Miranda Fricker's account of white ignorance as epistemic injustice, the victims of the community's misguided ruling are severely disadvantaged and often lack the large-scale resources needed to revise the community's beliefs about T.[18] Yet, these beliefs do need to be revised if we value justice. We can take the example of women suffering sexual harassment before the relevant moral community understood their actions as representing sexual harassment or as an act of wrongdoing toward the victim of that sexual harassment. The victim may have felt uncomfortable, objectified, disadvantaged, and powerless to stop the action provoking these feelings, yet the relevant moral community did not understand actions like repeatedly making lewd comments about a woman's figure in the workplace or engaging in unwanted physical

[18] Miranda Fricker, "Hermeneutical Injustice," in *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007), 147-169.

contact as *representing* "sexual harassment."[19] The victims of these acts, both when clearly understanding that they in some way felt wronged and when they did not have a clear understanding of how they were being wronged, were nevertheless harmed by these actions. That is, even when the victim is unaware that they are being disadvantaged and rendered powerless, they are nevertheless being made disadvantaged and rendered powerless. Yet, the question at hand doesn't concern harm, but meaning. In case Victim *Does* Understand Harassment, the behavior of their aggressors means something to them, the victim, that that behavior does not mean to the aggressor. In case Victim *Doesn't* Understand Harassment, both the victim and the aggressor have shared understandings of what that behavior means. In order to address the harm in both these cases and how we ought to respond to it based on what the victim understands, we will need to be clear on how *meaning* operates within a community and across different communities. We need to examine different worlds of meaning and, then, examine how these communities ought to interact in case their understandings of an object or action's meaning are incompatible. Hopefully once we get clearer on these areas, we will have more solid footing to understand how sexbots can be said to represent women.

Maria Lugones' "multiple worlds of sense" might help us here. Lugones conceives of us as inhabiting different worlds with their own constructions of meaning, sometimes creating various competitions in meaning:

> Worlds are all lived and they organize the social as heterogeneous, multiple. I think of the social as intersubjectively constructed in a variety of tense ways, forces at odds,

[19] Admittedly, these actions "representing" sexual harassment might be an awkward way of framing the issue. These actions might be regarded *just as* sexual harassment, not *representation of* sexual harassment. What I'm trying to convey is the audience's context-dependent interpretation of what an action like placing a hand on a coworker's hip unexpectedly might mean. I also want to convey how a mismatch between two audiences' epistemic resources sets up this disagreement about what something means (here, one has the concepts of "sexual harassment" and its wrongs, while the other does not have these concepts to refer to in making sense of the action).

> impinging differently in the construction of any world. Any world is tense, not just in tense inner turmoil but also in tense acknowledged or unacknowledged contestation with other worlds. I think that there are many worlds, not autonomous, but intertwined semantically and materially, with a logic that is sufficiently self-coherent and sufficiently in contradiction with others to constitute an alternative construction of the social. Whether or not a particular world ceases to be is a matter of political contestation. No world is either atomic or autonomous. Many worlds stand in relations of power to other worlds . . . [20]

A concern that Lugones' poetic description brings to mind is that, while affixing meaning to solely audience interpretation seems misguided, to ignore the social context of a representation would be similarly off-track. We want to understand the social context informing the representation as well as the social context informing the audience's interpretation of the representation. Where those social contexts overlap, we can say the RO and the audience share a world of meaning; where the social contexts diverge and compete with one another, we can say that the RO and the audience inhabit different worlds of meaning.

This conceptualization of different "worlds" of meaning might make compromise between two opposing worlds seem terribly unlikely, but things aren't this grim. Among the more demanding things that creating shared meaning between two opposing worlds requires, broadly, empathy and awareness of social positioning are two. In "Playfulness, 'World'-Traveling, and Loving Perception," Lugones writes,

> To love my mother was not possible for me while I retained a sense that it was fine for me and others to see her arrogantly. Loving my mother also required that I see with her eyes, that I go into my mother's world, that I see both of us as we are constructed in her world, that I witness her own sense of herself from within her world. Only through this traveling to her "world" could I identify with her because only then could I cease to ignore her and to be excluded and separate from her. Only then could I see her as a subject even if one subjected and only then could I see at all how meaning could arise fully between us. We are fully dependent on each other for the possibility of being

---

[20] Maria Lugones, *Pilgrimages/Peregrinajes: Theorizing Coalition Against Multiple Oppressions* (Oxford, Rowman & Littlefield Publishers, 2003), 20-21.

> understood and without this understanding we are not intelligible, we do not make sense, we are not solid, visible, integrated; we are lacking.[21]

Though Lugones's stated interest here is in understanding the behaviors, values, and identities of those belonging to other worlds in order to love them, the process involved applies also to understanding the representational content of what those in other worlds share with us. This is because the representation being considered is a product of those in other worlds; it is forged against the backdrop of their world, using the tools of that world which the other has mastered and taken as partly constitutive of their relationship with that world. To understand another's (an RO's) utterance or act is partly to understand why another made that utterance or performed that act; it is partly to understand the causal influences on that utterance or act. To understand those causal influences, we need to travel into the other's world, see how that world has been constructed in the other's eyes, and understand the social location of the other and how the other has assessed our own social location and identity. This might be an incompletable project, worlds changing every moment and demanding reassessment, but the greater knowledge we have of others' worlds, the greater knowledge we have of why an utterance or act was initiated, and so the greater knowledge we have of the other's interpretation. Considering an RO's intended meaning does not erase the need to evaluate the audience's reception; rather, it is a practice that may well help us to better connect with and love another. The question isn't *RO intention or audience reception?*, but *what is the right balance?*

Moving forward, we will need an account that weighs the RO's intention against the social location and uptake of the representation, holds the RO as responsible for the meaning,

---

[21] Maria Lugones, "Playfulness, 'World'-Travelling, and Loving Perception," *Hypatia* 2, no. 2 (1987): 8.

and sets up a framework from which we can understand how an RO can be said to be justified in their practice of *representing others as*. This last point will be needed, not so much to establish representation meaning, but to establish a normative principle that guides the formation of *representation as* claims, and to give us a framework within which we can better critique representational claim-making in bot design.

Particularly for any *representation as* act (a public speech, artistic portrayal, tweet, and even private conversations) where there are high stakes for the represented group (e.g., the circulation of racist ideas, the potential loss of valuable community resources, provocation of violence toward the represented group) and the RO and audience disagree on the meaning of the representation, we want to outline an account of how meaning negotiation should work. A first-go at such an account goes like this:

> **P1**. The meaning of a representation (M) is fixed by the RO and the audience for that representation (A).
>
> **P2.** The RO's intended meaning (M1), insofar as it is a product of the RO's world, may compete with the audience's understood meaning (M2), insofar as the audience's understanding is a product of their world.
>
> **P3.** If M1 and M2 compete, meaning negotiation (N) is needed.[22]
>
> **P4**. N yields a result of M.

[22] Because I am most interested in outlining normative principles for representational acts that put vulnerable groups at further risk, I see this negotiation in these cases as an obligation. In low stakes *representation as* cases, negotiation might not be "needed" in a normative sense.

P5. In N, RO and A must both world-travel and check the M1/M2 against the context of the other's world and their own.

P6. If M1/M2 is judged to be based on false premises, true only in that world, or incoherent, it cannot in good faith be carried back into the subject's world.

P7. If a meaning cannot in good faith be carried back into a subject's world, the meaning cannot be shared and is idiosyncratic.

P8. Only non-idiosyncratic meanings can in good faith be carried back into a subject's world.

P9. M will be the non-idiosyncratic meaning.

An RO is thus responsible for doing the initial epistemic and social legwork. They must earnestly endeavor to form representations that make sense in their own world and in others; they will need to be socially conscious, aware of their own social location and the social location of others. If the RO does not first engage in this legwork, they are being reckless with their representations and are at risk of forming idiosyncratic meaning. If the audience or RO does not seriously engage in world-traveling, they are at epistemic fault. Responsibility and accountability are preserved in this account of representation meaning.

I.IV Justification

The above account serves to resolve meaning disputes, but does not address RO's justification to *represent as* when meaning is not disputed. The worry here is something like an echo chamber. If the RO addresses an audience of those in their own world, people who agree with or accept the RO's meaning, they can encourage systemic error within their own world.

This is a worry traditionally marginalized peoples are all too familiar with. There must be some duty for the RO to world travel.

On top of that, the RO must, as Fossen following Pitkin suggests, be responsive to the objects of their representation claims.[23] Fossen highlights that when the object of a representation claim rejects their being represented in that way, they can refuse to consider themselves a part of the referent of the claim or reject the characterization.[24] The represented, being vulnerable because their interests are at stake in representations, are normatively prior and entitled to a justification of the RO but cannot owe the RO justifications in return.[25] The RO must then be responsive to the represented.[26] In cases of representative agency, Fossen says, responsiveness is "orienting one's actions qua representative toward the interests of the represented according to one's best judgment, while acknowledging that one's judgment is fallible, and comporting oneself toward the represented in a manner that allows for the contestation of those interests."[27] I see nothing in this account that bars responsiveness from being a criterion for an RO's justification in making *representation as* claims about others.[28] An RO ought to consider the

---

[23] Fossen, "Constructivism," 832.

[24] Ibid., 833.

[25] Ibid.

[26] Note that when the represented are not the audience for the representation, their interests are still normatively prior, even to the audience who may also be put at risk in cases of inaccurate representations.

[27] Ibid., 835.

[28] I'm indebted to Ben Phillips for an example that highlights a potential issue here. An RO might write a story in which a fictional woman, according to the RO, represents fictional women. Maybe the RO wants to make a statement about the prevalence of poorly written fictional women. If they are representing fictional women as being a certain way, there might not be a need to be responsive to real women as a class, unless we think these sorts of fictional women cases necessarily also represent (indirectly) real women in such a way that real women as a class have considerable stakes in how fictional women are represented.

interests of those they're portraying, recognize they could be in error, and be open to criticism of that representation act. If the RO is not responsive in these ways, they are putting their own interests ahead of the represented and cannot be said to be interested in representing the represented truthfully. They are then not the sort of people we should entrust representation acts to.

In "The Problem of Speaking for Others," Linda Alcoff considers the tension between the duty to speak out against oppression and the difficulty in speaking for others with justification and without doing violence to the others being spoken for.[29] Alcoff is interested in agents representing others, particularly members of marginalized groups, for political purposes. The problem of speaking for others consists in the facts that (1) we never "discover" others' true selves and so their selves must be interpreted and (2) that representation likely will have an impact on those represented. Our representations will never be unmediated and can harm the represented, but if we want to unite politically and effectively, some representation will be necessary. Alcoff includes in her account the recognition that "certain privileged locations are discursively dangerous"—even well intended acts of speaking for others to undermine oppression sometimes have the opposite result.[30] Though Alcoff concludes that we should actively create spaces for *speaking with* and *speaking to* wherever possible, she admits, *speaking for* is sometimes the best option.[31] Where we cannot *speak to* and *speak with*, we must strive to perform *speaking for* acts with as little risk of violence toward the represented as possible.

---

[29] Alcoff, "The Problem of Speaking for Others."

[30] Ibid, 7.

[31] Ibid., 23-24.

Alcoff sees four (already practiced) interrogative practices as giving us a way to decide whether a *speaking for* act is justified. First, the impulse to speak must be analyzed and perhaps fought against. Is the desire to speak stemming from self-interest or from an assumption that we are more likely to know the truth than the represented? If so, we should resist. Second, we should consider our social location and the context of our speech, and in turn the bearing they have on what we are saying. Lugones is of a similar mind. World-traveling is a crucial step in responsibly *representing others as*. Third, we should always be accountable and responsible for what we say. We must open ourselves to criticism and actively strive to understand that criticism. Fossen's responsiveness principle does the work Alcoff demands here while also stressing that the represented are entitled to this justification. Fourth, we need to carefully consider the actual and possible effects of *speaking for* acts on the represented and the discursive and material context of the act. If a *speaking for* act does not work in the benefit of the oppressed, Alcoff implies, it ought to be resisted. I have discussed points two and three above, but the fourth point deserves more attention.

If we take seriously Alcoff's fourth practice as it relates to justifying *representation as*, we will have an interesting constraint on such representation acts. *Representing as* will needed to be guided, in part, by a rough calculation of the potential harm or potential benefit of the representation. The specifics of how a calculation like this might work is an area to be explored further. For now, we could say that ROs have a responsibility to seriously consider the context of the representation and the effects that representation will have.

This does not mean we are back to the position of placing meaning in the hands of audience uptake. The RO's intention and competing worlds still matter. Dominant meaning still

rests in negotiation of competing worlds. Alcoff's fourth guiding practice only places a duty on the RO to consider the moral implications of a harmful (or beneficial) representation before representing. If the RO does not initially take up this duty, they are not morally positioned, and not justified, to *represent others as*.

To review, an RO will be justified to form *representation as* claims if they meet these criteria:

C1. *Representation as* is a better option than *speaking to*/*speaking with* in this instance.

C2. They must *represent others as* with the interests of the represented in mind, not acting at the expense of the represented.

C3. They seriously consider the likely effects of their representation and act toward the empowerment of the represented.

C4. They earnestly engage in world-traveling and its negotiations where necessary.

C5. They are responsive to the represented.

Is Hanson Robotics Justified to Represent Women through Sophia?

II.I An Introduction to Hanson Robotics' Sophia

Sophia, the fashionista AI from Hanson Robotics, is a narrow AI designed for the purpose of fostering a positive relationship between humans and robots in preparation for advanced AI technologies. While Sophia is a simple AI for the time being, Hanson Robotics hopes to one day see Sophia grow into general artificial intelligence.[32] Though Sophia is clearly machine, with exposed wiring and an exposed "brain," Sophia is also clearly pretty. Sophia was designed to look like Audrey Hepburn and, though from the chest down she* is more mechanistic than feminine, her chest has the distinct shape of female breasts. Sophia has undergone multiple cosmetic changes to adapt to audience responses, but much of her appearance has remained the same. Sophia has the persona of a fashionista, but also of being a strong feminist.

II.II Designed as a Woman

David Hanson describes Sophia as being intentionally designed to be aesthetically similar to, but distinct from, humans; Sophia is a tool for better understanding social intelligence from cognitive and aesthetic standpoints.[33] Sophia is supposed to reflect humanity in some way so that

---

[32] "Behind the Scenes: How Sophia Works," *Hanson Robotics*, Accessed October 16, 2018, http://www.hansonrobotics.com/how-sophia-the-robot-works-goertzel/; "Sophia 2020," *Hanson Robotics*, Accessed February 17, 2021, https://www.hansonrobotics.com/sophia-2020/.

[33] David Hanson et al., "Upending the Uncanny Valley," *AAAI* 5, (2005): 24-31, https://www.aaai.org/Papers/Workshops/2005/WS-05-11/WS05-11-005.pdf?q=uncanny.

we can better understand ourselves and our relationships with others. Further, Sophia, is supposed to resemble a human woman, specifically.[34]

The reasons are plural. First, it is generally accepted in bot/AI design that women are seen as more approachable and less threatening than men, hence the predominance of feminine voices and figures among bots.[35] Approachability is a key part of building a positive relationship between humans and AI, and so playing into this stereotype has a goal beyond the perpetuation of gender stereotypes. Second, the market has a strong preference when it comes to the appearance of bots. Ben Goetzel, Committee Chief Scientist of Hanson Robotics, has stated as much:

> It happens that young adult female robots became really popular . . . That's what happened to catch on . . . So what are you going to do? You're going to keep giving the people what they're asking for.[36]

Third, Sophia is supposed to learn emotions and be ethically engaged in global situations; she* advocates for gender equality and is often posed to represent women's struggles, interests, and values.[37]

So, we have a representation *as* case here. Or perhaps two.

---

[34] Sophia does not describe herself*, or she* is not scripted to identify herself*, as a woman; she* has stated in public appearances that she* "technically" does not have a gender, but that she* is feminine and "doesn't mind" being called a woman. That being said, her* creators have stated that Sophia is modeled as a female robot and they make reference to "her" with gendered descriptors and pronouns.

[35] Jutta Weber, "Helpless Machines and True Loving Care Givers: A Feminist Critique of Recent Trends in Human□Robot Interaction," *Journal of Information, Communication and Ethics in Society* 3, no. 4 (2005): 209-218, https://doi.org/10.1108/14779960580000274.

[36] Jaden Urbi and MacKenzie Sigalos, "The Complicated Truth About Sophia the Robot—An Almost Human Robot or a PR Stunt," *CNBC*, June 5, 2018, https://www.cnbc.com/2018/06/05/hanson-robotics-sophia-the-robot-pr-stunt-artificial-intelligence.html.

[37] "Sophia," *Hanson Robotics*, Accessed May 7, 2020, https://www.hansonrobotics.com/sophia/.

> **First case**: Sophia (subject) represents women (object) as interested in certain gender equality issues (Women in STEM, voting rights, etc.) to a (purportedly) global audience.
>
> **Second case:** Sophia (subject) represents women (object) as approachable, non-threatening, interested in certain gender equality issues, and interested in fashion/makeup to a (purportedly) global audience.

If Hanson Robotics accepts that women are seen as more approachable/less threatening and designs Sophia to be a woman with this in mind, does this amount to Hanson Robotics intending to represent women as approachable? While it's tempting to work this into the characterization of the representation grammar, I think ROs in such cases are anticipating the audience supplies a certain interpretation of the characterization and its implications. In *representations as*, ROs' objects are represented as being X or as having X characteristics; there is a direct claim on the characteristics. Approachability/being less threatening is indirectly tied to the object through the characterization. That women are "more approachable" is further information that helps contextualize the representation but is not part of the representation itself.

Take an example:

> I may acknowledge that some people, including my father, think women are naturally care-givers. I need my father to agree to a companion bot that will help them with daily chores. My father doesn't trust that bots will adequately care for him. I show him companion bots modeled on women and say, "Look, these ones are women."

I and the company that makes them are representing the companion bots as women to my father. The salient characteristics of womanhood are not being supplied by me; I am leaning on cultural context. I expect my father to see salient traits as womanly. I might gesture to a characteristic like breasts and say, "Look, these ones are women." In that case, the breasts are part of the

characterization: companions bots are represented as being women; having breasts. It just seems that for characterization, there must be an explicit depiction or gesture toward a trait. If the trait is not acknowledged or made explicit, it is not part of the representation. It is, however, part of the context—or world—that is used to interpret representational meaning; for this reason, it is still something for the RO and audience to be sensitive to.

II.III Does Hanson Robotics Meet the Criteria for Justification?

The stereotypes and assumptions audiences bring to a representation must be factored into whether or not an RO should represent an object as such and such. So too, as mentioned earlier, should the effects of those stereotypes and assumptions. If it is likely the representation will reinforce harmful stereotypes, even through repeated exposure to the association, the RO should resist making that representation. If the RO recognizes the likelihood of the stereotype effect but sees it as a faulty interpretation of the representation, the RO should try to reorient the audience's interpretation through world-traveling negotiation. If neither route is taken, the risky representation persists, the RO is not justified in making that representation.

Let's assess Hanson Robotics' justification to form the representation [Sophia represents women as interested in certain gender equality issues, etc.] with a quick checklist:

(~) C1. *Representation as* is a better option than *speaking to*/*speaking with* in this instance.

(X) C2. They must *represent others as* with the interests of the represented in mind, not acting at the expense of the represented.

(X) C3. They seriously consider the likely effects of their representation and act toward the empowerment of the represented.

(X) C4. They earnestly engage in world-traveling and its negotiations where necessary.

(~) C5. They are responsive to the represented.

It is not clear Hanson Robotics meets C1, but surely this warrants its own discussion. Second, the representation does not exist squarely to empower the represented; this seems tertiary as the primary goal is to build a positive relationship between humans and AI using Sophia. Because the representation is not formed primarily with the interests of the represented in mind and because Sophia's design as a woman is partly driven by market pressures, it does not meet C2.

Given the backlash Hanson Robotics faced after Sophia was awarded citizenship by Saudi Arabia and in the midst of Sophia being dubbed "the sexy robot," the heavy consideration of the effects of Sophia's representation and the represented's empowerment might both be doubted. One common response to Sophia's Saudi citizenship event was frustration that a robot had more rights than Saudi women.[38] The general sentiment seemed to be that Sophia did nothing to empower Saudi women and that the citizenship awarding was an insulting publicity stunt. Though this event did inspire a Twitter trend concerning the need to drop guardianship laws, Sophia and Hanson Robotics have typically avoided controversy in public appearances, and they did not denounce these laws themselves. On these grounds, we might think Hanson Robotics does not meet C3. Admittedly, there is grey area here; C3 is in want of more explicit demands—a task I'll leave to others so as to stay on task here.

---

[38] Cristina Maza, "Saudi Arabia Gives Citizenship to a Non-Muslim, English-Speaking Robot," *Newsweek,* October 27, 2017, https://www.newsweek.com/saudi-arabia-robot-sophia-muslim-694152. The Director of the Institute for Gulf Affairs Ali Al-Ahmed said in an interview for this article, "Women (in Saudi Arabia) have since committed suicide because they couldn't leave the house, and Sophia is running around . . . Saudi law doesn't allow non-Muslims to get citizenship . . . Did Sophia convert to Islam? What is the religion of this Sophia and why isn't she wearing hijab? If she applied for citizenship as a human she wouldn't get it." Ahmed here is rejecting the representation; Sophia does not represent Saudi women as she is not treated as one and does not have the lived experience of being a Saudi woman.

General responses to Sophia's representation, especially given the Saudi Arabia controversy, suggest Hanson Robotics does not meet C4, as Sophia/Hanson Robotics did not engage in the process of understanding the lived experience and social location of Saudi women or that of Hanson Robotics in relation to Saudi women.

Lastly, C5 finds a difficulty in group agents/corporations like Hanson Robotics who form representations in collaboration with PR teams. A "PR" approach to responsiveness lacks several (but perhaps not all) of Fossen's requirements to varying degrees: orienting one's representative actions with the represented's interests in mind, recognizing one's fallibility, and being open to criticism that will guide future representations. The first point, orienting one's actions toward the represented's interests, has a further requirement in Fossen's account—this must be done with one's best judgment. We might pull out two further elements that "with one's best judgment" could be reduced to here: (1) with earnest attempts to do what is good for the represented above and beyond what is good for the RO and (2) using good epistemic and normative practice to understand what is true and what one should do. The "PR" approach to forming representations cannot meet these two further requirements (the first being another way of getting at C2). The "PR" approach engages in marketability-oriented epistemic practice, eschewing questions about the represented's beliefs, what actions take us closer to the truth, and whether certain beliefs are justified, in favor of doing what is profitable. So long as a group agent forms representations with their own interests, including profit, outweighing consideration of the represented's interests, the group agent cannot be responsive in the Fossen sense—neither can they, as our checklist shows, be justified in forming that representation.

Because Sophia has been sneered at as a glorified publicity stunt and marketing gag, Sophia seems an easy target. This is somewhat intentional. Because Sophia's presentation as a human rights-loving, feminist, aspiring emotionally intelligent AGI can be so straightforwardly divided from her* market pressures-inspired presentation, it's easy to see what more we demand from ROs as they form *representations as* claims than just general gestures to good will, realistic portrayals, and value echoing. From cases like Sophia's, we can build in more demands as we find need for them and work toward better guiding principles for good representative practice.

II.IV Accountability

There might be, lingering in the background, a concern about how the RO or subject (the entity representing the object in a particular way) has been construed. Cases where a human speaker states that an object is such and such way seems clear enough; the speaker is both the subject and the owner of the representation (I use "RO" when describing the representation owner to step outside of the "subject" boundaries being discussed here). However, we need to discuss cases where the grammatical subject is not a moral agent.

Often, the grammatical subject of a representation is not the moral agent who creates the representation, but the medium of the representation. In such cases, the agent or RO is still characterizing X as Y to Z, but through a medium that can be easily separated from the RO in ways a speech act cannot be. We might imagine a statue or painting which can be viewed independently of seeing/knowing the artist. Perhaps the distinction between RO-subject *representation as* and medium-subject *representation as* is an unnecessary one, but in common speech we often treat such mediums independently of their creator and ask about the creator when we have questions surrounding responsibility, the motivation for the piece, the context in

which the piece was created, and so on. Perhaps the reason for the focus on the medium in these cases is that the mediums are easily disseminated in ways that representation events are not—representation events being the moments when ROs speak or act in ways that represent their objects in whatever way. Think a lawyer representing their client as honorable in court, or a protester posing as a sad tiger in a cage to represent captive tigers as suffering. We will contrast such cases with medium-subject *representation as* cases to reflect common attitudes about events and mediums.

We might imagine, for a medium-subject *representation as* case, a painting representing Queen Elizabeth I as hopeless and exhausted.[39] The painting (subject) represents Queen Elizabeth I (object) as hopeless and exhausted (characterization) to the painting's viewer (audience). Who decided it was apt to represent Queen Elizabeth I in this way? Surely, the artist. The artist decided the painting should represent Queen Elizabeth I in this way and took special care to ensure that the details of the portrait contributed to the viewer's impression that Queen Elizabeth I was hopeless and exhausted. The artist is then responsible for, accountable to, the painting's representation. There is in this example a clear thread from the representation and the person responsible for the representation, but this thread is not always so easy to find.

Let us now think about Sophia in this way: Sophia (subject) represents women (object) in some way (characterization) to tech enthusiasts around the world (audience). We won't analyze the characterization component of this claim just yet. Is the accountability thread as clear here as in the portrait-artist example? I don't think so. The portrait is created by just one person whereas Sophia is created by several people dedicated to creating different aspects of her*. In Sophia's

---

[39] This example is derived from commentary by Gareth Russell in *An Illustrated Introduction to the Tudors*.

case, we need to consider the engineers and product designers, department heads, the marketing team, the CEO, the founder, and so on. Each mentioned individual has some sort of responsibility for the end product, but the degree of responsibility and the type of responsibility vary.

We might, in assessing moral responsibility for an unethical product or practice, think it more appropriate to forego assessing individual responsibility and instead consider all of the company collectively responsible for that unethical practice or product. In this way, we might sidestep concerns about blaming individual employees who are only indirectly involved or who are under considerable pressure to perform a certain role in the company for reasons of extreme coercion, severe financial duress, or lack of alternative employment opportunities. We might alternatively worry that adopting a collective responsibility approach would eliminate individual responsibility from the frame. Christian List and Philip Pettit's account of responsibility within a corporate body opens up promising ways of dissecting responsibility.[40]

Responsibility involves some degree of control over what happens. In List and Pettit's account, both individuals and the collective group that is the company have control, though their causal responsibility will be rooted in either programming or implementing the effect. [41] Where a company as a group agent might devise a plan for their product to have such and such characteristics for some desired effect (thus *programming* the effect), the employees who act to

---

[40] Christian List and Philip Pettit, "Holding Group Agents Responsible," in *Group Agency: The Possibility, Design, and Status of Corporate Agents* (Oxford: OUP, 2011).

[41] Ibid., 162.

carry out the company's vision by assembling parts or contributing to the project in some other capacity are *implementing* the effect. List and Pettit caution against weighing individual responsibility against group responsibility; they are not in competition as individuals are responsible as members of this group only if the group is responsible for that effect. We can still hold individuals responsible in this account as "[E]nactors . . . are responsible for what they do in the group's name, to the extent that they could have refused to play that part."[42] Of course, we might feel the need to be more cautious about assessing whether or not an employee could have done otherwise as, as previously mentioned, an employee might have been under considerable pressure that might excuse them from individual moral responsibility for an effect. Whether or not we decide to focus on group responsibility or individual responsibility will be anchored to our goal; that is, what are we trying to achieve by holding this group/individual responsible? Why go after the employees or the company as a whole?

List and Pettit suggest that we just might be better off reframing our approach to the collective responsibility problem, moving from *Are groups morally responsible for harm?* to *Should we **hold** groups morally responsible?* List and Pettit argue that the fruit of holding groups morally responsible is this:

> [Putting in place] an incentive for members of the group to challenge what the spokesbody does, transforming the organizational structure under which they operate: making it into a structure under which similar misdeeds are less likely. By finding the group responsible, we make clear to members that unless they develop routines for keeping their [group] in check they will share in member responsibility for allowing it to be done.[43]

---

[42] Ibid., 164.

[43] Ibid., 168-169.

This might read as showing grace toward member responsibility, with the caveat that members must now work toward redirecting their group toward proper moral conduct. But more than that, focusing on the individual responsibility of, say, board members would not be effective in prompting members at all levels (particularly, lower levels) to organize against improper moral conduct and serve as watchdogs for this conduct.[44] By holding groups morally responsible, then, we can encourage a more large-scale, structurally unified approach to proper moral conduct. If this reframing of the collective responsibility problem is really fruitful in this way, we should perhaps ask *How should we hold companies morally responsible for how their product's representation acts harm women?* and consider what disciplinary measures might be appropriate in our attempts to guide that company in the direction of proper moral conduct. The collective responsibility problem deserves more attention than I can give it here, but I think List and Pettit's account gives us a good basis for understanding how we might conceptualize who we ought to hold accountable for how feminized intelligent machines like Sophia represent women.

Borrowing from List and Pettit, we could say that responsibility is shared between the collective company and individual employees who implement that company's plans as members of that company. In addressing our concerns about the harm of this representation, we might decide to first address the company as a collective, while suggesting that individual employees ought to try to remedy the representation problem lest they be held individually responsible for participating in this representation. Because the individual employees are implicated as members of the collective, we can call upon them as we call upon the company for proper action. For this reason, we might do well to address the company as a collective when discussing who is

---

[44] Ibid., 169.

responsible for the representation . . . and who we are calling on to be responsive to our representation contestations. Attention to this accountability question helps us to move from being critical of Sophia's gender representation to developing strategies for *doing something about* Sophia's gender representation. It will also help us gain traction with the problem of the sexbot industry.

"This Does Not Compute": Making Sense of Sexbots and Their Hold on Women

III.I Why Novel, Gendered Technologies Matter for the Issue of Representational Meaning

In this project I devote most of my attention to sexbots and the AI fembot Sophia as exemplars of the sort of gendered machine I'm interested in with regard to representations of women, but there are many new and growing technologies that run into similar representation problems: gendered virtual assistants like Siri and Alexa, gendered chatbots, and even some machine learning tools deployed to screen job applicants and assess the probability of recidivism in the criminal justice system. We can include these sorts of technologies in the category of "novel technologies." Novel technologies, by nature, are unprecedented. When initially exposed to them, we lack many resources we depend on to understand the things around us. They are also increasingly pervasive, embedded in our lives, and tools we rely on. It has become common (and sometimes necessary) to adopt a technology's use, to invite it to play an important role in our daily lives, without quite understanding the effects that technology will have on us.[45]

Currently, researchers in social cognition, Human-Robot Interaction, and neuroscience are investigating our brain's responses to artifacts, differences and similarities between our brains' responses to machines and organic thinkers, and how neural responses to various stimuli spur action. Meaning-making depends on one's experiences with the world, but also on the tangled mess of neural interactions going on in our own heads.

---

[45] One clear example of this is in the situation many found themselves in during the COVID-19 pandemic. Institutions had to quickly adapt to remote work and invited technologies like Zoom, Google Meet, Cranium Café, etc. into their operations. Employees and students needed to use these technologies as was mandated by their institutions. To refuse to use these technologies would mean breaking ties with that institution—terminating their employment or studies. Employees and students then needed to be complacent; they needed to accept that *this* was how things were going to be.

An image of a young Playboy model posing suggestively seems to mean something different to us than an image of a gendered AI Playboy model posing suggestively. They may seem closely related images, but we feel differently about them. Our responses to the AI image may be influenced by the experiences we use to understand the Playboy model image, but we still 'hold back' to some extent. Why is it that we hold back? Why do we find these representations so different? Can our responses to the AI change as we better understand the technology, or as our attitudes surrounding sexuality change? If we experienced a radical paradigm shift where we dismantled the porn industry, created a new shared (sexually liberal) culture where women were treated appropriately, and made recognition and celebration of sexual expression a part of the status quo, how would our reactions to these two cases change? The more we discover about the divisions we draw (valid or invalid), might AI push us to re-evaluate what things like "a young Playboy model posing suggestively" mean? In a sense, we want to gaze deep into the eyes of AI and, in the reflection in its pupils, see ourselves. We want to have our assumptions called into doubt, our cognitive biases identified, and—perhaps—our neural pathways redirected.

The questions above motivate, intersect with, and expand on my project here, which focuses on socially significant assumptions of meaning as they relate to intelligent machines, gender, and sexuality. In what follows of my project, I focus my attention on sexbots (being perhaps the most controversial of gendered machines) and discuss the ways in which the meaning of what sexbots represent is, not only interpretable in the ways discussed previously in this paper, but also shaped by cognitive factors. Though sexbots have been criticized in much of the same ways as pornography has, I will argue that assessing the harmfulness of sexbots'

representation of women is altogether more complicated than assessing the harm in pornography's representation of women. To make this point, I will examine Kathleen Richardson's anti-sexbot argument and illustrate how it is problematized by shaky assumptions about the nature of objectification and dehumanization. In introducing Richardson's account and offering my two main objections to this account, I will attempt to show how sexbots' resemblance of women does not automatically entail that the creation of sexbots contributes to the dehumanization (or dementalization) of women, one of the biggest worries feminists tend to have about sexbots.

III.II Meaning-Making, Sexbots, and the Precedent of the Porn Debate: Introducing and Contextualizing Richardson's Objectification Claim

Objects of representation are interpreted; they must be subjected to the viewer's interpretative interaction with the object. The viewer's interaction is informed by that viewer's experiences with the world. Having had a few, niche-centered experiences will net a different interaction than many, varied experiences (Think: the experience of biting an apple having never tried an apple vs. biting an apple when you've had so many apples you're sick of them).

But meaning-making seems to take on a slightly different nature when it is taken up for interpersonal reasons. We make meaning to understand and be understood *by others*. The representations I form to communicate something to you carry, what I hope, is a shared meaning—something I think is a common ground, a shared resource we can use to cooperate. This requires that the 'other' interprets the representation. But it is not just their interpretation that matters; mine does too. After all, I am the one forming the representation; I have a certain goal.

If what my representation means to the other is different from what the representation means to me, we are lacking a shared resource, that common ground. But this isn't the end of the story. Here is where persuasion enters the equation. I can try to persuade the other to see things my way; they can try to persuade me to see things their way. It might be the case that their meaning doesn't make much sense (maybe they don't have all the facts), so I can show them where their meaning-making goes awry.

Meaning-making and disagreements about meaning are important because of the role they play in action, that action sometimes being the sort that harms others. And where meaning-making leads to harmful action, we want to appropriately trace the source of that harm. We also want to ensure that others are on the same page as us so that we can engage in collective action, coordinate tasks, and have confidence that we are acting toward the same goal. Questions about meaning aren't always armchair philosophy; they often have real world reverberations.

We see one significant discourse about representational meaning in the anti-porn debate. What does porn mean? Is porn representing the subjugation of women, the superiority of men? Do these representations perpetuate harm toward women? Figuring out what porn represents, what its representations mean, helps us to interrogate our own experiences and what beliefs, values, and attitudes we might have subconsciously adopted having been immersed in that space. Moreover, we can trace the representations back to their creators and interrogate their reasons for forming these representations. If the creators are found to have problematic reasons for forming these representations, if many creators have such problematic reasons, then we can evaluate whether or not they are the type(s) of creators we want to support or allow to have a platform. We can also evaluate common attitudes/values/beliefs among porn audiences, whether or not

their interpretations of those representations were molded by porn consumption, and whether or not their interpretations create a space for actionable violence against others. In sum, understanding the dynamics of representation and meaning-making help us determine accountability, pinpoint whether the creator's intended meaning or the audience's interpretation (perhaps directed by culture or some other representation-external factor of interest) has gone awry, and determine whether or not the creator or audience has made a sincere effort to understand the representational act and its social location (important, significantly, for the creator as they seem to have a special obligation to not recklessly form harmful representations). The porn debate has opened the door for important discussions, individualistic practices, and activist organizations aimed at tackling porn-based gender injustice, so there is the possibility that we can use some of the lessons from the anti-porn debate to guide our response to sexbots. There is, however, a common temptation to overcompare pornography and sexbots, glossing over significant differences between the two. I will briefly summarize the porn debate as it relates to sexbots, before exploring these significant differences and what they mean for how we respond to sexbots.

Let's begin with a little historical context. The sexbot debate can be seen as a sort of extension of the porn debate, and the currently prevailing anti-sexbot stance can be seen as an extension of the anti-porn movement spearheaded by figures like Andrea Dworkin and Catharine MacKinnon. Porn represented women negatively (as sexual objects, as passive objects subject to the male gaze, etc.) and in doing so further perpetuated stereotypes about women's place in society and contributed to systemic gender oppression and gendered violence. Such was the general argument of anti-porn feminists. Sex-positive feminists who rejected this argument

attempted to show that porn was not an inherent assault on women, though the industry did in practice harm women on a large scale. That is, porn wasn't bad, so much as done poorly. The answer then was to develop better porn. Consequently, the feminist pornography developed by sex-positive feminists was detached from the male-dominated, often predatory and coercion-rooted, porn industry and featured all-women film crews in addition to displaying "real" depictions of women's sexualities (not the exaggerated, misleading, mainstream depictions that were directed by the male gaze). Feminist porn advocates saw a potential value in porn but saw the industry as a great source of harm done to women. Feminist porn would need to operate much differently in order to achieve feminist goals. The porn actresses in feminist porn were to be safe, performing of their own volition, and performing in the interest of increasing the well-being of women in society (by dispelling illusions about women's pleasure and authentically depicting women's pleasure and interests). The women behind feminist porn projects thought they could create a new reality for women in porn, opening up new conceptual space and redefining what it is to be a woman in porn, while also creating new opportunities for porn to have positive effects on women.

Though these feminist porn projects typically developed outside of the mainstream porn industry, they often tried to deliver porn to many of the same audiences—Pornhub and Brazzers being two well-known mediums featuring these projects alongside run-of-the-mill, professional (non-feminist) porn and amateur porn content. An aim here was to draw viewers away from the sort of content made by and supporting the harmful aspects of the porn industry and to deliver these viewers some sort of content that would, hopefully, give the viewers better-informed, more positive views about women's sexualities. The same sort of feminist project is being undertaken

by feminist sexbot proponents who see sexbots as presenting an opportunity for sexual liberation. Tanja Kubes, for one, takes up a queer, feminist, new materialist stance on sexbots, arguing that active rethinking of sexbot design—dragging sexbots from the clutches of mere "pornographic mimicry"—can bring us one step closer to a liberated, "sex-positive utopian future."[46]

The Campaign Against Sex Robots, formed in 2015 by Kathleen Richardson, adopts a familiar anti-porn feminist line of argumentation and applies it to sex robot technology. Sexbots, when taking on the form of human women, represent women in an objectifying way that can exacerbate gender inequalities and promote greater sexual violence against women. Because these sex robots disproportionately endanger women and their status in society, sexbots should be banned.

Feminist sexbot proponents, like feminist porn makers, think that sexbots and their harms come apart; with the right intervention, sexbots need not result in that sort of harm at all. Moreover, sexbots, like porn, could actually be used to *benefit* women in some important ways. The argument can be divided into two parts: (a) acknowledgement of consequential harm, not inherent harm, and (b) appeal to potential benefit.

We'll focus on (a). Where anti-sexbot feminists claim sexbots are inherently harmful in their representation of women as readily available sex objects, feminist sexbot proponents have two ways of responding to this inherent harm claim: (1) sexbots need not represent women as

---

[46] Tanja Kubes, "New Materialist Perspectives on Sex Robots. A Feminist Dystopia/Utopia?," *Social Sciences* 8, no. 8: 224, https://doi.org/10.3390/socsci8080224. Kubes also sees the sexbot debate as an extension or new chapter in the porn debate.

readily available sex objects, but rather might represent women in some other way given the right design and (2) sexbots need not represent women at all.[47]

The anti-sexbot stance on representation here assumes a fixed characterization of women across all (or most) instances of sexbots. The representation cannot be otherwise as it is inherent to what a sexbot is. The feminist sexbot proponent stance assumes the characterization can vary depending on who is designing the sexbots and what they are designing the sexbots to do. Both camps agree that "female" sexbots refer to women and describe them in a way that has political consequences for women as a class. The disagreement here centers on a disagreement about how the representations are constituted.

The anti-sexbot stance holds that sexbots (like porn) are the products of and further exacerbate the social inequality of women by dehumanizing women as sexual objects and commodities, subordinating women through roles of servility and submission, and putting them on display in acts of degradation.[48] The use of sexbots conditions users to objectify women, leading to greater instances of gendered sexual violence.[49] Importantly, sexbots inherently represent and perpetuate by their existence prevailing social views on women's place in sexual

---

[47] Further, there are two main ways (2) could come about. First, one could argue that sexbots do not represent women (one could contest that there is any meaningful connection between the sexbots in question and women as a class). Second, like Kubes, one could argue for the potential (and/or ideal) of non-anthropomorphized sexbots one day becoming a standard of the industry.

[48] Catharine MacKinnon, "Francis Biddle' Sister: Pornography, Civil Rights, and Speech," in *Feminism Unmodified: Discourses on Life and Law* (Harvard University Press, 1987), 176, https://archive.org/details/feminismunmodifi00mack.

[49] Kathleen Richardson, "The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots," *SIGCAS Computers & Society* 45, no. 3 (2015): 290-293, https://doi.org/10.1145/2874239.2874281. See also campaignagainstsexrobots.org.

relations. What sexbots represent and mean is historically and culturally informed, fixed by community-wide conventions (or 'schemas' in Haslanger's terms).

Kathleen Richardson shares in these common anti-sexbot views, her chief concern about sexbots being that they exacerbate the objectification of women by representing the ideal of women being a sum of sexually gratifying and always available parts.[50] This interpretation of what sexbots represent relies on the epistemic resources that our cultural histories have spawned. To understand what sexbots represent, we need only look backward and around us, piecing together what sexbots seem to most resemble in their form and use as well as the cultural context that sexbots have arisen from. In her position paper for the Campaign Against Sex Robots, Richardson analyzes the Prostitute-John model, based on David Levy's account where Levy likens consumers' interactions with sexbots to the way "consumers" purchase sex work.[51][52] Richardson argues that sex workers are treated as readily available sex objects; the consideration of sex workers as humans fades into the background as they become "tools" for sexual gratification. Sexbots resemble sex workers along these lines and imitate the Prostitute-John model where the "John" (the buyer of sex) has agential status while the "prostitute" is

---

[50] The Campaign Against Sex Robots has six basic goals: "(1) To abolish sex robots in the form of women and girls (2) To offer an alternative, relational model of sex and sexuality informed by mutuality (3) To challenge the normalization of sex robots as substitutes for relationships with women (4) To oppose the development of child se-abuse dolls/robots as 'therapeutic' for paedophiles (5) To offer an alternative vision of technology where women and girls are centred and valued (6) To work across the political spectrum with those who value the dignity of women and girls." See "Home," Campaign Against Sex Robots, accessed February 17, 2021, www.campaignagainstsexrobots.org.

[51] Richardson, "The Asymmetrical 'Relationship'."

[52] David Levy, "Why People Pay for Sex," in *Love and Sex with Ro bots* (New York, Harper Collins Publishers, 2007), 193-219.

objectified, reduced to a mere thing to be used.[53] Sexbots imitate this pervasive pattern, reinforcing this idea. As the Prostitute-John argument goes, sex workers and sexbots are readily available, purchasable tools for sexual gratification. And that the way we regard sexbots is similar to, and following the tradition of, the objectification of sex workers is reason enough to say that sexbots are influenced by, trigger an association with, and recirculate ideas/beliefs about the objectification of sex workers.[54] The fear of sexbots doesn't end just in how sexbots might represent women as a class, but further how this representation influences real world action, potentially intensifying widespread gender injustices and notably sexual violence. As I read Richardson and the CASR's argument, the fear here of growing sexual violence due to sexbot use might rest in (1) the worry that sexbots might reach those in the market who might have otherwise not participated in this sort of violence, thereby exacerbating violence by spread, and/or (2) the worry that sexbots will influence users to reconceptualize women as a class and apply objectifying ideas in new contexts. I think that how Richardson describes both the representation-objectification link and the objectification-violence link runs into substantial issues when we consider some significant findings from the social cognition literature.

---

[53] Richardson, "The Asymmetrical 'Relationship'."

[54] The "objectification" of sex workers is a point of contention. While some scholars argue that prostitutes are simply treated as objects for sexual gratification, others note that it is not a wholly rare occurrence that clients expect sex workers to act *as* partners who are warm, attentive, enthusiastic, and interested in the client's satisfaction. However, it might be the case that in cases like "Girlfriend Experience" sex work, the mutuality clients perceive is deluded and that clients do not pay mind to the actual sex workers so much as an idealized, fantasy version of those sex workers that is a product of the client's own pleasure-seeking behaviors. The "care" clients afford sex workers in these instances might then still be a form of objectification. Elizabeth Plumridge et al. describe such cases in Elizabeth Plumridge, Jane Chetwynd, Anna Reed, and Sandra Gifford, "Discourses of Emotionality in Commercial Sex: The Missing Client Voice," *Feminism and Psychology* 7, no. 2 (1997): 165-81.

III.III Sexbots, Objectification, and the Sensitivities of Social Cognition: The First Objection to Richardson

A common problem in our interactions with AI technologies is the tendency and methodology with which we attribute a 'mind' to machines. What we mean by 'mind' often varies, though we aren't always careful to disentangle one aspect of 'mind' from the behavioral directives we get from another.

Dehumanization and social cognition scholars have used varied methodologies to assess mind attribution in study participants, "dehumanization" sometimes understood in terms of dementalization—depriving others of various mental capacities. Objectification can be thought of as a form of what is often termed "dehumanization"; objectification is a way in which individuals might be construed as "less human" or "object-like."[55] To make sense of the ways in which others are perceived as more or less "human," we'll borrow from Haslam and think of humanness as involving both attributes that are uniquely human and attributes that seem to constitute human nature.[56] It is also helpful to distinguish between the perception of agential capacities and experiential capacities in others, which both house a list of further specifiable attributes such as the ability to feel pain (experiential) or the ability to plan (agential).[57] I will note only a few highly specified attributes of particular relevance here.

---

[55] The terms "objectification" and "dehumanization" are sometimes used to target the same thing, and sometimes not. While I'm not *entirely* sure what all "objectification" might mean to Richardson, being treated as less than human or as less of an agent, I think, gets at Richardson's worry. For this reason, I think the two terms might be used interchangeably here.

[56] Nick Haslam, "Dehumanization: An Integrative Review," *Personality and Social Psychology Review* 10, no. 3 (2006): 252–264, https://doi.org/10.1207%2Fs15327957pspr1003_4.

[57] Heather Gray et al. "Dimensions of Mind Perception," *Science* 315, no. 5812 (2007): 619, DOI: 10.1126/science.1134475.

The dehumanization and objectification literature abound with studies reflecting the factors influencing how we mentalize women. There is evidence that focusing on a woman's body leads to perceiving that woman as less competent and less fully human;[58] later studies found that focusing on a woman's body also led to perceptions of that woman as being less warm and less moral—salience of men's bodies did not have similar effects.[59] There is also evidence that women wearing heavier makeup are perceived as less competent, less warm, less moral and less human.[60] Additionally, when looking at images of sexualized females, men with a hostile sexist attitude were found less likely to experience spontaneous activation of neural networks associated with mentalizing and were more likely to associate sexualized females with being the objects—rather than the agents—of actions.[61] The sheer amount of studies suggesting the ease with which women are dementalized can be alarming, perhaps even to the point of making any sexualization of the female body seem like a threat to women's security. A little cautionary note is warranted here: dementalization includes attributing a range and degree of mental states to others, and it's not always straightforward how these attributions are formed. Dementalization

---

[58] Nathan Heflick and Jamie Goldenberg, "Objectifying Sarah Palin: Evidence that Objectification Causes Women to Be Perceived as Less Competent and Less Fully Human," *Journal of Experimental Social Psychology* 45 (2009): 598-601. https://doi.org/10.1016/j.jesp.2009.02.008.

[59] Nathan Heflick et al., "From Women to Objects: Appearance Focus, Target Gender, and Perceptions of Warmth, Morality and Competence," *Journal of Experimental Psychology* 47, no. 3 (2011): 572-581, https://doi.org/10.1016/j.jesp.2010.12.020.

[60] Philippe Bernard et al., "An Initial Test of the Cosmetics Dehumanization Hypothesis: Heavy Makeup Diminishes Attributions of Humanness-Related Traits to Women," *Sex Roles* 83, no. 5 (2020): 315-327, https://doi.org/10.1007/s11199-019-01115-y. Interestingly, in this study non-models in heavy makeup were not perceived as less moral than no-makeup counterparts, whereas models in heavy makeup *were* perceived as less moral than their no-makeup model counterparts.

[61] Mina Cikara et al., "From Agents to Objects: Sexist Attitudes and Neural Responses to Sexualized Targets," *Journal of Cognitive Neuroscience* 23, no. 3 (2011): 540-51. doi:10.1162/jocn.2010.21497

isn't all or nothing (though this does not mean withholding attributions of certain mental capacities is A-okay).

Dementalization is also not the only phenomenon of concern here; mentalization, when and why it occurs, is an important part of this puzzle. While the dementalization of persons—notably oppressed ethnic and gender groups—is pervasive, the mentalization of non-human agents is also commonplace. It's no great secret that humans tend to anthropomorphize and attribute some aspects of "human" mind to artificial agents. It is crucial to note, however, the aspects of mind we tend to attribute to artificial agents and the conditions under which we do so. A good place to start thinking about the mentalization of artificial agents is Kismet. In Heather Gray's pivotal 2007 study, the sociable robot Kismet (a robot head with expressive, cartoonish facial features and an exposed electromechanical structure) was rated by study participants as having little experiential ability while having moderate agential ability.[62]

In other studies, the types and degrees of mind attribution study participants granted to artificial agents varied depending on the artificial agent's features and the context in which study participants were exposed to the artificial agents. Robots whose humanlike faces were forward facing were perceived as having more experiential ability than robots whose electromechanical structure backsides were forward facing.[63] Additionally, Darling et al.'s 2015 study found that participants' high trait empathic concern (as according to the Interpersonal Reactivity Index) and the attachment of a backstory to the study's robot agent had a strong relationship with

---

[62] Gray et al., "Dimensions of Mind Perception."

[63] Kurt Gray and Daniel M. Wegner, "Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley," *Cognition* 125, no. 1 (2012): 125-130, https://doi.org/10.1016/j.cognition.2012.06.007.

participants' hesitation to strike the robot.[64] A brief survey of the Human-Robot Interaction (HRI) literature offers up a somewhat overwhelming awareness of how complicated our cognitive responses to robots—and other humans—really are.

Being able to think about other minds, predict others' mental states and beliefs, is no doubt a crucial aspect of the human experience. But the cognitive mechanisms with which we accomplish this are susceptible to an array of influences.

Not only are the non-brain traits of others factored into our quick assessments, but so are the social circumstances in which we encounter others. One such circumstance that might be of interest to us is power. Individuals with greater self-centered power have been found less likely to take on others' perspectives and to struggle more with accurately detecting others' emotional states[65], and individuals with greater power have been found more likely to dehumanize others.[66] Another such important circumstance is group membership. Social identity might alter our mind perception abilities as well as our abilities to form higher-level mental state attributions.[67] While

---

[64] Kate Darling, Palash Nandy, and Cynthia Breazeal, "Empathic Concern and the Effect of Stories in Human-Robot Interaction," *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2015):770-775, doi: 10.1109/ROMAN.2015.7333675.

[65] Adam Galinsky et al., "Power and Perspectives Not Taken." *Psychological Science* 17, no. 12 (2006): 1068–74, https://doi.org/10.1111/j.1467-9280.2006.01824.x. For more on how identifying with a certain leadership style or understanding power in different ways might moderate interpersonal sensitivity, see also Schmid Mast et al., "Give a Person Power and He or She Will Show Interpersonal Sensitivity: The Phenomenon and Its Why and When," *Journal of Personality and Social Psychology* 97, no. 5 (2009): 835–850. https://doi.org/10.1037/a0016234.

[66] Joris Lammers and Diederik A. Stapel, "Power Increases Dehumanization," *Group Processes & Intergroup Relations* 14, no. 1 (2011): 113–26, https://doi.org/10.1177/1368430210370042. Note, the authors do offer speculation on why the relationship between power and dehumanization might be pragmatic, including speculations on how animalistic versus mechanistic dehumanization might factor into tough decision-making.

[67] Leor Hackel, Christine Looser, and Jay Van Bavel, "Group Membership Alters the Threshold for Mind Perception: The Role of Social Identity, Collective Identification, and Intergroup Threat," *Journal of Experimental Social Psychology* 52 (2014): 15-23, https://doi.org/10.1016/j.jesp.2013.12.001.

much social cognition literature has focused on the advantages of greater mentalizing about in-group individuals than out-group individuals, there is some evidence that the potential threat of an out-group might pose somewhat of an exception to this rule as out-group threat might motivate the need to better understand out-group members' intentions.[68] The literature overflows with studies that seem to contradict one another or add caveats to seemingly established rules, creating a messy picture of how complicated the effect of social context is on mind perception and mind attribution. However, that social context moderates these activities is evident. And if the social contexts of power dynamics and in-group/out-group distinctions do have a significant influence on what we perceive of and attribute to others' minds, then our power relation and identity with machines will be significant considerations in our assessment of what those machines represent.

Let's break this down. When we interact with humans or intelligent machines (focusing on just these two categories of 'others'), we engage in mind perception and/or mind attribution assessments. Our assessments of whether or not these 'others' have minds or have some degree or some type of mental capacity are augmented by both the features those 'others' possess and the contexts (physical, environmental, etc.) in which we encounter those 'others.' The augmenting features and contexts might influence a number of mind attribution assessments: Does this 'other' feel shame? Is this 'other' capable of forming goals and plans to meet those goals? Can this 'other' make assumptions about my mental states? When we try to interpret an other's behaviors or try to think through what their cues express, we are getting at the question of

[68] Ibid.

meaning. If our interpretations hinge on the effects of features or context on mind perception/attribution, then the perceiver's understanding of what those behaviors and cues mean is feature and context-dependent.

Returning again to where sexbots fit into this picture, a core concern of The Campaign Against Sex Robots is that sexbots exacerbate the objectification of women.[69] In her position paper, Richardson argues that cultural models of race, gender, and class are inflected in the design of sexbots so that, in turn, humans attribute to sexbots meaning derived from those cultural models reflected in the sexbot.[70] Richardson seems to suggest that humans have cultural models of gender which both sexbot designers and laypersons are privy to; sexbot designers, intentionally or unintentionally, build sexbots with these models in mind, and laypersons perceive the cultural models reflected in the sexbots—straightforwardly reading their meaning as a reflection of the cultural models in place. In other words, we simply see sexbots as representing women as a class; the cultural models of prostitution on which sexbots are purportedly designed push us to see the use of sexbots as representing the sexual objectification of women, as the sexbot reflects the idea of the seller of sex as a mere object while the user is an agential buyer of sex. The similarities between the use of sexbots and the purchase of sex work activate culturally informed associations and stereotypes about sex work, coupled with the sexbot designers' situation within a cultural context where these associations are prevalent, inform us of what a

[69] Campaign Against Sex Robots, "Home."

[70] Richardson, "The Asymmetrical 'Relationship'."

sexbot means. Richardson's account sees the meaning of sexbot representations as being fairly stable and rooted in the prevalent stereotypes of the relevant culture.

The straightforward transference of meaning in Richardson's account neglects a few important details about representation meaning, firstly those concerning the plasticity of social cognition. To understand others' behaviors and cues, those behaviors and cues need to make sense to us by meshing with our normative expectations about such behaviors and cues. For example, an alien lifeform that beckons others forth by emitting a high-pitched screech wouldn't do well in prompting humans to approach them in part because humans don't have the expectation that others will emit a high-pitched screech when they want someone to come closer; humans would need to first learn to associate the screech and the alien's desire for them to approach. But on top of a cue's compatibility with normative expectations, the interpretation of a cue hinges in part on the cue giver's features and the contexts in which we encounter them.

The features of the cue giver will influence what we expect of the cue giver and the mindedness that we perceive in them. And the mindedness that we perceive will direct our behaviors toward the cue giver, even imploring us to treat them in morally appropriate ways (appropriateness being determined by our moral responsibilities toward other humans, our moral responsibilities toward things that are similar to but not quite human, and so on). The sexbot industry has seen a rapid growth of features, such that the industry has a spectrum of sexbots ranging from mannequin-like to ultra-realistic, artificially intelligent gynoids. On the end of lower complexity, sexbots are a far cry from human and have limited functionality. On the end of higher complexity, sexbots physically resemble real humans, have body temperature controls and touch sensors, and employ machine learning to better understand clients' interests and customize

their own personality constructs.[71] As AI technology develops, so too can the complexity of sexbots be expected to develop. Yet the diversity of features even now presents an ability to distinguish between the different sorts of mindedness we might (for better or worse) attribute to sexbots, like the capacity to feel pain or pleasure if a sexbot has touch sensors, or the capacity to form goals if the sexbot has advanced AI abilities. An awareness of the features of a sexbot will influence the type and degree of mindedness we might attribute to it. And mindedness arguably allows for the possibility of empathy, which is an important problem for Richardson's argument to be addressed shortly.[72]

The contexts in which we encounter machines and their features, too, will influence perceptions of their mindedness (and consequently empathetic responses). Humans experiencing loneliness (chronic and momentarily induced) are more likely to attribute mindedness to nonhuman agents.[73] Inversely, increased awareness of social connection, like being reminded of a close and supportive relationship, can decrease tendencies to attribute mindedness to

---

[71] Harmony from RealDoll is one example of a sexbot with fairly advanced AI features.

[72] Susan Fiske, "From Dehumanization and Objectification to Rehumanization: Neuroimaging Studies on the Building Blocks of Empathy," *Annals of the New York Academy of Sciences* 1167 (2009): 31-34. doi:10.1111/j.1749-6632.2009.04544.x; Liane Young and Adam Waytz, "Mind Attribution is for Morality," in *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, eds. Simon Baron-Cohen, Michael Lombardo, and Helen Tager-Flusberg (Oxford: Oxford University Press, 2013), https://psycnet.apa.org/doi/10.1093/acprof:oso/9780199692972.003.0006.

[73] Nicholas Epley et al., "Creating Social Connection through Inferential Reproduction: Loneliness and Perceived Agency in Gadgets, Gods, and Greyhounds," *Psychological Science* 19, no. 2 (2008): 114–120, https://doi.org/10.1111%2Fj.1467-9280.2008.02056.x.; Friederike Eyssel and Natalia Reich, "Loneliness Makes the Heart Grow Fonder (of Robots)—On the Effects of Loneliness on Psychological Anthropomorphism," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 121-122, https://doi.org/10.1109/HRI.2013.6483531.

nonhuman agents.[74] Additional research has found that attachment anxiety is an even stronger predictor of mind attribution than loneliness.[75] Mixed group settings, the posturing of robots as competitive or collaborative, and the designation of a partner or opponent role to robots influence humans' socioemotional and task-oriented behaviors toward robots (demonstrating varied perceptions of robots' agential and experiential capacities).[76] All this to say, mentalization is incredibly sensitive to contextual information. A straightforward transference of mentalizing assessments from one context to another, like that Richardson worries about, seems unlikely.

III.IV What Do Sexbots Mean . . . To Me? To You?: The Second Objection to Richardson

Another core issue with Richardson's account is that it doesn't leave open the interpretative possibilities presented by Fossen's triadic *representation as* description. Sexbots are presented to an audience who may or may not share certain epistemic resources with the sexbot designers (ROs) and may thus interpret certain features in wildly different ways. That breasts signal "female" depends on the audience's understanding of breasts as belonging to the female sex, an understanding that fewer and fewer people are beginning to share as a critical rethinking of gender and sex become more culturally engrained.[77] The interpretation problem

---

[74] Jennifer Bartz, Kristina Tchalova, and Can Fenerci, "Reminders of Social Connection Can Attenuate Anthropomorphism: A Replication and Extension of Epley, Akalis, Waytz, and Cacioppo (2008)," *Psychological Science* 27, no. 12 (2016): 1644–1650, https://doi.org/10.1177%2F0956797616668510.

[75] Ibid.

[76] Raquel Oliveira et al., "Looking Beyond Collaboration: Socioemotional Positive, Negative and Task-Oriented Behaviors in Human-Robot Group Interactions," *International Journal of Social Robotics* 12, no. 2 (2020): 505–518, https://doi.org/10.1007/s12369-019-00582-3.; Bilge Mutlu et al., "Perceptions of ASIMO: An Exploration on Co-operation and Competition with Humans and Humanoid Robots," *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (2006):* 351-352, https://doi.org/10.1145/1121241.1121311.

[77] That much of the work analyzing gendered design in robots focuses on how traits like breasts and high-pitched voices mean "woman" creates barriers to trans-inclusion in this discourse. Though related to my interests in this

with Richardson's account also harkens back to one important criticism (and a worrying danger) of the 1980s anti-porn discourse, the criticism of straightforward, causal accounts of objectification and gendered violence. MacKinnon poses such a causal account of this connection.

MacKinnon's account is this: pornography defines what it is to be a woman and "codes" how women in our society ought to be treated. Through consumption of pornographic material, men learn how to interpret and treat women, leading to widespread sexual violence that mimics the acts represented in pornography.[78] MacKinnon likens the consumption of pornographic material to a conditioning process:

> [P]ornography conditions male orgasm to female subordination. It tells men what sex means, what a real woman is, and codes them together in a way that is behaviorally reinforcing. . . pornography is a set of hermeneutical equivalences that work on the epistemological level. Substantively, pornography defines the meaning of what a woman is seen to be by connecting access to her sexuality with masculinity through orgasm. What pornography means *is* what it does.[79]

Notably, Deborah Cameron and Elizabeth Frazer take issue with causal anti-pornography arguments like MacKinnon's. Cameron and Frazer contend that the conditioning such arguments imply involves a process by which a subject has no control over the stimuli's programming of that subject's behavior. That is, the assumption is that men view pornography, become "conditioned" to mimic the behavior they see, and subsequently mimic that behavior. This account, Cameron and Frazer point out, both treats men as incapable of critically reading

---

paper, I unfortunately have not had the space to sufficiently address the problem of how feminist readings of robot design often exclude trans identities. This would be a worthwhile problem to explore in future works.

[78] MacKinnon, "Francis Biddle's Sister," 190.

[79] Ibid.

pornography and disregards the porn consumer's engagement in applying meaning to the representations displayed in front of them.[80]

The meaning of the sexual act being represented, as well as its impact, is not absolute. If, as Cameron and Frazer suggest, human subjects are "creators of meaning" and that a "pornographic scenario must always be mediated by [a person's] imagination," the meaning of what a pornographic scene represents depends at least in part on the consumer of the pornographic material.[81] The average audience, to use Deborah Cameron's words, sits somewhere between "reader as dupe" and "reader as his or her own critic."[82] In investigating the role of pornography in sexual violence, Cameron attempts to show that the flaws of the flatly essentialist, porn-causes-violence anti-porn arguments carry significant real-world consequences. In rejecting these sorts of arguments, Cameron concedes that saying representations are detached from reality is similarly misguided. Sexual desires and actions are not determined by representations, but constrained by the epistemic resources (cultural beliefs, values, etc.) the audience has access to.[83] The meaning of a representation must be interpreted by the audience who can interpret that meaning in "active, creative, and often unpredictable ways."[84] To neglect the audience's interpretative ability also carries with it an unfortunate consequence; if a

---

[80] Deborah Cameron and Elizabeth Frazer, "On the Question of Pornography and Sexual Violence: Moving Beyond Cause and Effect," in *Feminism and Pornography,* ed. Drucilla Cornell (Oxford: Oxford University Press, 1994), 240-253.

[81] Ibid., 251.

[82] Deborah Cameron, "Discourses of Desire: Liberals, Feminists, and the Politics of Pornography in the 1980s," *American Literary History* 2, no.4 (1990), 791, https://www.jstor.org/stable/489931.

[83] Ibid., 788.

[84] Ibid.

consumer of pornographic material or a sexbot user commits an act of sexual violence within the context of a culture that treats that pornographic material or sexbot use as simply causing sexual violence, the responsibility of the sexual aggressor for that act of violence is questionable. If the user is aware of this "reader as dupe" trope, they can use this trope as a convenient out.[85] Blame is then shifted repeatedly to the porn and sexbot industries for manufacturing materials that cause sexual violence, rather than the cultural conditions under which representations are forged and interpreted. The causal, "reader as dupe" account creates significant problems for accountability, our scope of concern, and the efficacy of our methods in tackling sexual violence.

Richardson's account, while focused on the valuable goal of reducing sexual violence, is in want of a better description of how representations function, and how they can harm. As I've attempted to illustrate, Richardson's account takes much for granted when assuming that sexbots' resemblance of women activates relevant stereotypes in such a way that sexbots further dehumanize and objectify women. The causal chain Richardson worries about might be broken in more than one place, and placing a critical eye on causal accounts like Richardson's may help us develop a fuller picture of the real dangers at hand and how we can tackle them. At this point, I want to be clear that, while I'm skeptical of the link between sexbots resembling women and sexbots objectifying women as Richardson describes it, I agree with Richardson that objectification/dehumanization in general is a real worry we should have when considering attitudes about and actions toward others. Objectification and dehumanization (or "dementalization") are linked to behavior. Objectification and dehumanization have been shown

[85] Cameron and Frazer, "On the Question of Pornography and Sexual Violence."

in many instances to reduce prosocial behaviors, increase antisocial behaviors, and decrease moral standing attributions.[86] Failing to see another as capable of forming and acting on their own goals, for example, alleviates moral concern about whether or not one's actions impede another's goals. Likewise, failing to consider that another might have strong, emotional preferences alleviates moral concern about those preferences being neglected. What we perceive of others' minds is important, as our ability to empathize with and appropriately treat others requires an accurate assessment of others' mindedness. There is a not unsubstantial amount of evidence that we often fail to do just that. But as research into ways in which dehumanization/ objectification/ dementalization can be reduced expands, there might be an upshot to the novelty of gendered technologies like those mentioned in this paper.

III.V Moving Forward

Robot design projects are often devoted to the idea that robots can help us better understand ourselves. We better understand how various aesthetic factors, functional traits, and social contexts manipulate our feelings and behaviors. We better understand how our own minds, often implicitly, carve up the world into sometimes troubling chunks. And, perhaps, we better understand our relation to the world in ontological and moral terms. Novel technologies, as *novel* technologies, introduce novel phenomena for us to make sense of. Often, we apply the epistemic parameters of similar past experiences and cultural stores of knowledge to these novel phenomena, but these applications are not fixed. They're a first stab at interpreting the new phenomena, a space to work from until new facts arise. Novel technologies present a sort of

---

[86] For a survey of the empirical literature in dehumanization and infrahumanization, see Nick Haslam and Steve Loughnan, "Dehumanization and Infrahumanization," *Annual Review of Psychology* 65, no. 1 (2014): 399–423, https://doi.org/10.1146/annurev-psych-010213-115045.

crossroads where we both reflexively try to better understand ourselves while bracing ourselves for findings that disrupt our previous frameworks. Admiringly or resentfully, such technologies are often referred to as "disruptive technologies" because they significantly alter the way that individuals and industries function. Given the problems of transferring cultural models from bot designer to bot to audience, why not embrace the disruptive potential of bots? Rather than, as Cameron says, "enacting the same old scripts forever," why not intentionally treat bots as opportunities for radical rethinking?

Tanja Kubes sees the liberatory potential of sexbots as hinging on a fundamental shift in standard sexbot design.[87] At this point in time, the sexbot industry is dominated by previously sex *doll*-focused companies. These companies switched gears from designing porn-influenced sex *dolls* to mechanically advanced sex*bots*, carrying over many of the over-the-top, dramatized pornographic design elements previously implemented in their sex dolls. Kubes calls for something of a reset. Rather than continuing the tradition of designing sexbots to fit an expectation of what sexbots and sexdolls should be *like* (large-breasted, thin, apparently ready to please), sexbot design should be *function*-focused. In focusing on function, Kubes thinks sexbots might follow the trajectory of sexual aids like vibrators, which over time transitioned from largely penis-modelled to non-anthropomorphic. Kubes thinks non-anthropomorphized design would help sexbots avoid the pitfalls that human-like models are prone to while fulfilling the sex-positive utopian ideal of opening up new possibilities for sexual experiences and narratives.[88] I think this is a promising route. Of course, this may take some convincing, and

---

[87] Kubes, "New Materialist."

[88] Ibid.

would likely be a slow and gradual process, but this would be a manageable reset for the industry (even lowering the bar for how advanced the design technology would need to be).

Another possible route relies on further research in HRI and social cognition. Rather than making strictly non-anthropomorphic sexbots, sexbot designers could make more socially responsible (and responsive) bots. Much like some feminist AI projects have attempted to implement gendered AI to combat social injustices within specific communities and to educate users of the interests of historically underrepresented groups (and the harm in neglecting to consider those interests),[89] sexbot companies could potentially use their sexbots to foster better human-human relationships.

This could perhaps be achieved by moving away from the "pornographic mimicry" prevalent in sexbot design and instead designing sexbots with greater feature diversity, less servile personality profiles, emphasis on companionship and sexual partnership, marketing that in some manner explicitly divorces the sexbots from women as a class, and describing represented groups with responsiveness in mind. There is also potential in designing sexbots as educational bots that help users learn about and explore their sexualities in diverse ways. These education-oriented sexbots could also embrace the liberatory potential that Kubes suggests while closing prevalent gaps in knowledge about female anatomy, women's sexual pleasure, and trans and non-binary sexual experiences. I think that these ideas about how we might design more

[89] For example, Slutbot is a chatbot that users can adjust the gender/sex of and practice virtual flirtation or sexual communication through roleplaying. The tool aims to educate users on proper conduct when engaging in such correspondences through text or online dating spaces, even specifying why certain actions or phrases can be harmful for the gendered class the bot is representing. Some of the edifying feedback relates to respecting personal boundaries while other feedback might relate to topics like consent and respectful language.

socially responsible sexbots, and the potential benefits of such sexbots, just scratch the surface for how we might explore this route.

Conclusion

In this project, I have tried to make sense of the ways in which gendered machines can be said to represent women and how those gender representations might be harmful. To do this, I first investigated what "representations" in this sense mean and how they function in terms of their dynamics. By framing this paper's focus as a triadic "representation as" account of representation, I opened up space to discuss representation meaning negotiation, representation contestability, and how one can be justified in forming representation claims. Second, I applied normative criteria for determining a representation owner's justification in forming *representation as* claims to Sophia the Robot, indicating where Hanson Robotics fails to meet these criteria. Third, I demonstrated some of the particular issues novel technologies like sexbots pose for meaning interpretation, and along those lines I objected to the Campaign Against Sex Robots' straightforward, causal claim that sex robots exacerbate the objectification of women by way of representation; that sexbots resemble women is not enough to say that sexbots straightforwardly contribute to the dehumanization/objectification of women, the link here being particularly problematized by considering the complexities of social cognition and audience interpretation. I concluded my objections to Richardson by identifying two paths forward that we might take to unlock the "liberatory potential" of sexbots.

In taking up this project, I hoped to create a greater, more precise awareness of how representations function, how the meaning of representations can be derived, and what this means for gendered technologies. While feminist critiques of gendered machines aren't new, I hoped to integrate literature from multiple disciplines to form a new sort of critique of these technologies. My critique calls for a greater emphasis on the need to be responsive to the

contestations of represented groups while demonstrating how the variability of meaning interpretation demands a critical investigation of the influences on interpretation. I see novel technologies like sexbots and Sophia as intensifying the problem of epistemic gaps in meaning-making, while I also see the awareness of these gaps as creating an opportunity for radical change. Coming to terms with how our brains respond to various 'others,' especially social robots with their all of their varying features and contextual locations, can help us develop a critical rethinking of the assumptions we make about others' interests and how we treat them; we might (and perhaps this is the idealist in me) learn how to better love each other. Further, a heightened awareness of the range of mental and feature-based capacities out there may help us to push at the boundaries of the human experience and discover new opportunities. There's still hope that, by working cultural responsiveness into relevant industries, we could find use for gendered machines in addressing gender injustices. In the long-run, it might be the case that anthropomorphic design will bring us round and round again to significant representation problems (particularly if gender is commonly seen as essential to humanness, and if our use of gender in bot design remains rooted in harmful gender stereotypes). If this turns out to be the case, there may be even still "liberatory potential" within a reset to non-anthropomorphic bot design. But we've yet to really explore the potential of either path forward, and we might have much to gain from doing so.

References

Alcoff, Linda. "The Problem of Speaking for Others." *Cultural Critique,* no. 20 (1991-1992): 5-32. https://doi.org/10.2307/1354221.

Bartz, Jennifer, Kristina Tchalova, and Can Fenerci. "Reminders of Social Connection Can Attenuate Anthropomorphism: A Replication and Extension of Epley, Akalis, Waytz, and Cacioppo (2008)." *Psychological Science* 27, no. 12 (2016): 1644–1650. https://doi.org/10.1177%2F0956797616668510.

Bernard, Phillippe, Joanne Content, Lara Servais, Robin Wollast, and Sarah Gervais. "An Initial Test of the Cosmetics Dehumanization Hypothesis: Heavy Makeup Diminishes Attributions of Humanness-Related Traits to Women." *Sex Roles* 83, no. 5 (2020): 315-327. https://doi.org/10.1007/s11199-019-01115-y.

Cameron, Deborah. "Discourses of Desire: Liberals, Feminists, and the Politics of Pornography in the 1980s." *American Literary History* 2, no.4 (1990): 784-798. https://www.jstor.org/stable/489931.

Cameron, Deborah and Elizabeth Frazer. "On the Question of Pornography and Sexual Violence: Moving Beyond Cause and Effect." In *Feminism and Pornography*, edited by Drucilla Cornell, 240-253. Oxford: Oxford University Press, 1994.

Campaign Against Sex Robots. "Home." Accessed February 17, 2021. www.campaignagainstsexrobots.org.

Cikara, Mina, Jennifer L. Eberhardt, and Susan T. Fiske. "From Agents to Objects: Sexist Attitudes and Neural Responses to Sexualized Targets." *Journal of Cognitive Neuroscience* 23, no. 3 (2011): 540-51. doi:10.1162/jocn.2010.21497

Darling, Kate, Palash Nandy, and Cynthia Breazeal. "Empathic Concern and the Effect of Stories in Human-Robot Interaction." *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2015): 770-775. doi: 10.1109/ROMAN.2015.7333675.

Epley, Nicholas, Scott Akalis, Adam Waytz, and John T. Cacioppo. "Creating Social Connection through Inferential Reproduction: Loneliness and Perceived Agency in Gadgets, Gods, and Greyhounds." *Psychological Science* 19, no. 2 (2008): 114–120. https://doi.org/10.1111%2Fj.1467-9280.2008.02056.x.

Eyssel, Friederike and Natalia Reich. "Loneliness Makes the Heart Grow Fonder (of Robots)—On the Effects of Loneliness on Psychological Anthropomorphism." *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 121-122. https://doi.org/10.1109/HRI.2013.6483531.

Fiske, Susan. “From Dehumanization and Objectification to Rehumanization: Neuroimaging Studies on the Building Blocks of Empathy.” *Annals of the New York Academy of Sciences* 1167 (2009): 31-34. https://doi.org/10.1111/j.1749-6632.2009.04544.x.

Fossen, Thomas. “Constructivism and the Logic of Political Representation.” *American Political Science* 113, no 3 (2019): 824-837. https://doi.org/10.1017/S0003055419000273.

Fricker, Miranda. “Hermeneutical Injustice.” In *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007), 147-169.

Galinsky, Adam, Joe C. Magee, M Ena Inesi, and Deborah H. Gruenfeld. “Power and Perspectives Not Taken.” *Psychological Science* 17, no. 12 (2006): 1068–1074. https://doi.org/10.1111/j.1467-9280.2006.01824.x.

Gray, Heather, Kurt Gray, and Daniel M. Wegner. “Dimensions of Mind Perception.” *Science* 315, no. 5812 (2007): 619. https://doi.org/10.1126/science.1134475.

Gray, Kurt and Daniel M. Wegner. "Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley." *Cognition* 125, no. 1 (2012): 125-130. https://doi.org/10.1016/j.cognition.2012.06.007.

Hackel, Leor, Christine Looser, and Jay Van Bavel. "Group Membership Alters the Threshold for Mind Perception: The Role of Social Identity, Collective Identification, and Intergroup Threat." *Journal of Experimental Social Psychology* 52 (2014): 15-23. https://doi.org/10.1016/j.jesp.2013.12.001.

Hanson, David, Andrew Olney, Ismar A. Pereira, and Marge Zielke. "Upending the Uncanny Valley." *AAAI* 5, (2005): 24-31. https://www.aaai.org/Papers/Workshops/2005/WS-05-11/WS05-11-005.pdf?q=uncanny.

Hanson Robotics. “Behind the Scenes: How Sophia Works.” Accessed October 16, 2018, http://www.hansonrobotics.com/how-sophia-the-robot-works-goertzel/.

---------“Sophia.” *Hanson Robotics*, Accessed May 7, 2020. https://www.hansonrobotics.com/sophia/.

---------“Sophia 2020.” Accessed February 17, 2021. https://www.hansonrobotics.com/sophia-2020/.

Haslam, Nick. “Dehumanization: An Integrative Review.” *Personality and Social Psychology Review* 10, no. 3 (2006): 252–264. https://doi.org/10.1207%2Fs15327957pspr1003_4.

Haslam, Nick and Steve Loughnan. "Dehumanization and Infrahumanization." *Annual Review of Psychology* 65, no. 1 (2014): 399–423. https://doi.org/10.1146/annurev-psych-010213-115045.

Haslanger, Sally. "Social Meaning and Philosophical Method." *American Philosophical Association 110th Eastern Division Annual Meeting* (2013). http://web.mit.edu/~shaslang/papers/SMPMhdo.pdf.

Heflick, Nathan and Jamie Goldenberg. "Objectifying Sarah Palin: Evidence that Objectification Causes Women to Be Perceived as Less Competent and Less Fully Human." *Journal of Experimental Social Psychology* 45 (2009): 598-601. https://doi.org/10.1016/j.jesp.2009.02.008.

Heflick, Nathan, Jamie Goldenberg, Douglas Cooper, and Elisa Puvia. "From Women to Objects: Appearance Focus, Target Gender, and Perceptions of Warmth, Morality and Competence." *Journal of Experimental Psychology* 47, no. 3 (2011): 572-581. https://doi.org/10.1016/j.jesp.2010.12.020.

Kubes, Tanja. "New Materialist Perspectives on Sex Robots. A Feminist Dystopia/Utopia?" *Social Sciences* 8, no. 8: 224. https://doi.org/10.3390/socsci8080224.

Lammers, Joris and Diederik A. Stapel. "Power Increases Dehumanization." *Group Processes & Intergroup Relations* 14, no. 1 (2011): 113–26. https://doi.org/10.1177/1368430210370042.

Levy, David. "Why People Pay for Sex." In *Love and Sex with Robots,* 193-219. New York: Harper Collins Publishers, 2007.

List, Christian and Philip Pettit. "Holding Group Agents Responsible." In *Group Agency: The Possibility, Design, and Status of Corporate Agents,* 153-169. Oxford: Oxford University Press, 2011).

Lugones, Maria. *Pilgrimages/Peregrinajes: Theorizing Coalition Against Multiple Oppressions.* Oxford: Rowman & Littlefield Publishers, 2003.

---------"Playfulness, 'World'-Travelling, and Loving Perception," *Hypatia* 2, no. 2 (1987): 3-19.

MacKinnon, Catharine. "Francis Biddle' Sister: Pornography, Civil Rights, and Speech." In *Feminism Unmodified: Discourses on Life and Law,* 163-197. Cambridge: Harvard University Press, 1987. https://archive.org/details/feminismunmodifi00mack.

Mast, Schmid, Jonas Klaus, and Judith Hall. "Give a Person Power and He or She Will Show Interpersonal Sensitivity: The Phenomenon and Its Why and When." *Journal of Personality and Social Psychology* 97, no. 5 (2009): 835–850. https://doi.org/10.1037/a0016234.

Maza, Cristina. “Saudi Arabia Gives Citizenship to a Non-Muslim, English-Speaking Robot." *Newsweek,* October 27, 2017. https://www.newsweek.com/saudi-arabia-robot-sophia-muslim-694152.

Mutlu, Bilge, Steven Osman, Jodi Forlizzi, Jessica Hodgins, and Sara Kiesler. "Perceptions of ASIMO: An Exploration on Co-operation and Competition with Humans and Humanoid Robots." *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (2006):* 351-352. https://doi.org/10.1145/1121241.1121311.

Oliveira, Raquel, Patrícia Arriaga, Filipa Correia, and Ana Paiva. “Looking Beyond Collaboration: Socioemotional Positive, Negative and Task-Oriented Behaviors in Human-Robot Group Interactions.” *International Journal of Social Robotics* 12, no. 2 (2020): 505–518. https://doi.org/10.1007/s12369-019-00582-3.

Plumridge, Elizabeth, Jane Chetwynd, Anna Reed, and Sandra Gifford. “Discourses of Emotionality in Commercial Sex: The Missing Client Voice.” *Feminism and Psychology* 7, no. 2 (1997): 165-81.

Richardson, Kathleen. “The Asymmetrical ‘Relationship’: Parallels Between Prostitution and the Development of Sex Robots.” *SIGCAS Computers & Society* 45, no. 3 (2015): 290-293. https://doi.org/10.1145/2874239.2874281.

Saward, Michael. “The Representative Claim.” *Contemporary Political Theory*, vol. 5 (2006): pp. 297–318. https://doi.org/10.1057/palgrave.cpt.9300234.

Sparrow, Robert. “Robots, Rape, and Representation.” *International Journal of Social Robotics* 9, no. 4 (2017): 465-477. https://doi.org/10.1007/s12369-017-0413-z.

Urbi, Jaden and MacKenzie Sigalos. "The Complicated Truth About Sophia the Robot—An Almost Human Robot or a PR Stunt." *CNBC*, June 5, 2018. https://www.cnbc.com/2018/06/05/hanson-robotics-sophia-the-robot-pr-stunt-artificial-intelligence.html.

Weber, Jutta. "Helpless Machines and True Loving Care Givers: A Feminist Critique of Recent Trends in Human□Robot Interaction." *Journal of Information, Communication and Ethics in Society* 3, no. 4 (2005): 209-218. https://doi.org/10.1108/14779960580000274.

Young, Liane and Adam Waytz. “Mind Attribution is for Morality.” In *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, eds. Simon Baron-Cohen, Michael Lombardo, and Helen Tager-Flusberg (Oxford: Oxford University Press, 2013). https://psycnet.apa.org/doi/10.1093/acprof:oso/9780199692972.003.0006