Social Good Versus Robot Well-Being: On the Principle of Procreative Beneficence and Robot Gendering

Ryan Blake Jackson[1] and Tom Williams[1]

***Abstract*— The principle of procreative beneficence (PPB) is the idea that parents should use all available genetic, reproductive, and other technologies to select the child, of the possible children they could have, who is expected to have the best life based on all available information. The application of this principle to human reproduction has been extremely controversial, but scholars have argued that it may be applied more aptly and less problematically to the creation of robots with significant moral status (that is, robots with some meaningful personal well-being). We argue that, while some arguments against the PPB in human reproduction are less relevant to robot production, the PPB is still fundamentally at odds with the broader social good when applied to the creation of robots.**

I. INTRODUCTION

Robot ethicist John Danaher has argued compellingly for ethical behaviorism in defining and assessing the moral status of robots [1]. By "moral status" Danaher means the degree to which an entity or its interests morally matter for the entity's own sake (also understood as considerability or moral patiency). Ethical behaviorism holds that a sufficient reason for believing that we have duties and responsibilities toward other entities, or that they have rights against us, (i.e., moral status) can be found in their observable relations and reactions to their environment and to ourselves. In other words, the ethical behaviorist perspective is that robots have significant moral status if they are roughly performatively equivalent to other entities that have significant moral status. For example, if a robot consistently behaves like it can feel pain, and if the capacity to feel pain is a ground of moral status, then the robot should be granted the same moral status as any other entity to which we ascribe moral status based on the capacity to feel pain.

Danaher also argues that eventually creating a robot with significant moral status is likely unavoidable. The performative threshold for some form of significant moral status may be quite low. Many people now accept that animals with limited behavioral repertoires (like chickens or mice) have *some* moral status in that they have welfare and that they should not be harmed unnecessarily or with undue cruelty. It seems feasible for robots of the near future to be (at least) behaviorally equivalent to such animals in all morally relevant aspects [1]. Furthermore, a low performative threshold may be desirable based on arguments that the moral risk of over-inclusivity in the circle of moral concern is lower than the moral risk of under-inclusivity. Based on a low performative threshold for moral status, it might be difficult to create robots that do not have some significant moral status. As Danaher argues, "it may require creating robots that lack any behavioral manifestation of intention, desire, or agency." [1]. Likewise, the human drive to create robots that cross the performative threshold for significant moral status may prove too overwhelming for any system of norms (legal or moral) to constrain.

A natural consequence of the fact that, unlike humans, robots with moral status will only be created intentionally through deliberative design processes is the idea that any duties and responsibilities towards robots extend into the design process, rather than only applying during the robots' "life". This idea brings Danaher to the field of procreative ethics, and specifically to the principle of procreative beneficence (PPB), which we describe in the next section. Danaher articulates several compelling reasons why the PPB, which many scholars view as immoral when applied to human procreation, may be more defensible when applied to the creation of robots with significant moral status. We argue here that the PPB is nonetheless fundamentally at odds with the broader social good, and thus immoral, even when applied only to the creation of robots.

II. THE PRINCIPLE OF PROCREATIVE BENEFICENCE

The interpretation of the PPB advanced by Danaher holds that, although one is not under any obligation to procreate (or create a robot with significant moral status), if one decides to procreate (or make such a robot) one is under a duty to procreate a child (or robot) with the best possible existence given current knowledge, technology, etc. This idea may seem unobjectionable at first glance, as parents naturally want the best for their children. However, the implications of applying this principle to human procreation given technologies like abortion, in vitro fertilization (IVF), and preimplantation genetic diagnosis (PGD) are extensive and troubling. The creator of the PPB, Savulescu, deals specifically with these technologies in its prescriptions, arguing that reproducers should select for genes in their offspring that are expected to lead to the best possible life or the most possible advantages [2]. The argument is straightforward in, for example, selecting between two otherwise identical embryos where one has a lower risk of eventually developing cancer. However, Savulescu also argues that the PPB mandates reproductive decisions that select for the "best" non-disease genes as well, where non-disease genes include those that cause or predispose a person to some physical or psychological state which is not a disease state, "e.g.

[1]Department of Computer Science, Colorado School of Mines, Golden, CO, USA {rbjackso, twilliams}@mines.edu

height, intelligence, character (not in the subnormal range)" [2]. Genes that yield even a weak probabilistic relationship to intelligence, criminality, sexuality, sex, and gender are genes that should influence decisions according to the PPB. Indeed, Savulescu gives intelligence, criminality, and sex as examples, and states that "[the PPB] implies couples should employ genetic tests for non-disease traits in selecting which child to bring into existence and that we should allow selection for non-disease genes in some cases even if this maintains or increases social inequality." Perhaps because of the obvious unpalatability of these implications, Savulescu emphasizes an individualistic and libertarian non-coerciveness in the proposed moral obligations of the PPB, despite the role that coercion and consequence routinely play in enforcing other moral oughts (including the idea that people should not smoke, which Savulescu cites as a moral norm that uses "should" in a manner comparable to his own meaning in "[people] should choose the best child").

The PPB has been controversial, and there exist many counterarguments to its application to humans [3], [4], [5], [6]. Among these, Danaher draws particular attention to the ideas that (1) the PPB "favours a welfarist test for the 'best life' that ignores or overlooks other aspects of the good life," (2) that epistemic limits on genetic diagnosis preclude one from identifying in advance which child will have the best life, and (3) that the PPB places too high a burden on potential procreators, particularly those that will carry and birth children [1].

Another counterargument is the idea that allowing some individuals to deliberately select for the "best babies" may result in negative externalities for the rest of society that produce a net harm which is fundamentally at odds with the consequentialist perspective presupposed by the PPB [6]. Specifically, Sparrow argues that, because prospects for well-being are a product of an interaction between an individual's social environment and their genetic phenotype, and because changing or selecting for genes is likely easier for individual parents than changing society, the PPB obligates parents to create children that represent the dominant social group or that can pass as members of privileged groups; in the US, for example, the PPB would have all parents producing tall, white, male, and heterosexual children to the fullest extent possible via available genetic technologies because existing systems of oppression and minoritization would decrease children's potential for well-being otherwise [6]. Proponents of the PPB emphasize that nobody is directly harmed on an individual level by making these choices (those who are never born and never exist cannot be harmed) [2]. However, considered in aggregate, the collective choice of many procreators to select for offspring according to prevailing bigotries would likely exacerbate those bigotries, thus harming existing minoritized individuals as well as future individuals born without genetic selection/enhancement. Further, though the tendency towards homogenization in accordance with current prevailing bigotries might not be *prima facie* immoral insofar as it could reduce the number of individuals who suffer under bigotry, it is fundamentally at odds with the libertarian non-coercion that proponents of the PPB rely on to differentiate it from more traditional eugenics. Even supposing that *de jure* coercion by the state is somehow avoided in a society that views genetic selection/enhancement as a moral obligation and broad social good, and views those who forego it as drains on shared resources and welfare systems, social and economic forces can be just as coercive. The PPB would morally obligate its adherents to engage in persuasive social campaigns to influence other's reproductive decisions, and market forces may eventually make it untenable to avoid genetic selection/enhancement of offspring since suboptimal offspring would be unable to effectively compete in a competitive (capitalist) world [6].

Arguing comprehensively against (or, hypothetically, for) applying the PPB to humans is outside the central focus of this paper and outside the areas of expertise of these authors, though we intuitively oppose any eugenicist project, individualized and non-coercive or otherwise. We thus turn our attention back to robots, and to Danaher's argument that the PPB actually may apply to robots more so than to humans because "the most compelling objections to the application of the PPB to human procreation ... carry much less weight when applied to robotic procreation". This is certainly true for some of the arguments against the PPB in humans. First, the burden placed on creators of robots by the PPB is much more reasonable and less problematically gendered than the burden placed on creators of humans. Second, the decision to create a robot is entirely voluntary in all plausible circumstances (barring coercive economic forces, but one is still extremely unlikely to accidentally create a robot). Third, one has much more control in creating a robot than in creating a human, and many of the epistemic limitations that apply to human procreation do not apply to robot creation; there can be more certain and fine-grained control over a robot's development and quality of life than a person's [1].

However, we mentioned above some closely related arguments against the PPB that we believe apply to the creation of robots just as much as to the creation of humans. These are the consequentialist argument about the PPB creating a net harm in a population regardless of the benefit to one's offspring, and the argument that the PPB would mandate compliance with prevailing bigotries, that its commitment to non-coercion seems improbable given social and economic forces, and that reproductive coercion stemming from immoral bigotries should be avoided for many reasons, not the least of which that it would counter efforts towards a just, equitable, heterogeneous, and diverse society that we view as morally desirable. We use robot gender presentation as a quintessential example to argue that the PPB can be fundamentally at odds with the broader social good in terms of creating robots with certain socially constructed identity attributes and significant moral status.

III. CREATING GENDERED ROBOTS

Robots do not have gender identities in the same way that humans do, but humans have a powerful natural tendency to ascribe gender to social robots based on various gender cues

that the robots present. Even non-anthropomorphic machines with minimal gender cues (e.g., only voice) cause gender-based stereotypic responses in humans [7]. Visual cues as simple as hair length can also cause gendering of robots and stereotypic responses [8]. While some designers have attempted to avoid gendering their artificial entities (e.g., the ostensibly genderless voice "Q" [9]), it remains to be seen whether robot designers can (or should) prevent ascriptions of gender to social robots.

Perceptions of robot gender affect the behaviors and judgments of human interactants. Research has shown that robot gendering influences measures of robot friendliness [7], knowledge [7], agenticness [8], communality [8], suitability for certain tasks/roles [8], [10], [11], persuasive capacity [12], trustworthiness [12], credibility [12], and likeability [12], [13]. Therefore, robot gender presentation will impact the experience and well-being of any robot with significant moral status. Thus, the PPB applies to the design of robot gender presentation just as much as it applies to any other robot attribute.

In a patriarchal society (like ours in the US) where women face certain gendered obstacles, subordination, and dangers that men do not [14], the PPB could be interpreted to prescribe that we give all robots with moral status male gender cues, especially because, in many cases, it is trivially easy for robot designers to manipulate their robots' gender presentations. Indeed, Savulescu makes essentially the same argument with respect to humans, stating "in a country [where] women are severely discriminated against... Procreative Beneficence implies that couples should test for sex, and should choose males as they are expected to have better lives in this society, even if this reinforces the discrimination against women." [2].

In response to the criticism that this practice would exacerbate the struggles of existing and future women, Savulescu presents two arguments. The first is that "it is unlikely selection on a scale that contributes to inequality would promote well-being. Imagine that 50% of the population choose to select boys. This would result in three boys to every one girl. The life of a male in such a society would be intolerable." [2]. We disagree with this argument for several reasons. First, reproductive selection for male offspring has already contributed to inequality in multiple societies [15]. Apparently parents in those societies determined that the benefits that maleness would confer to their offspring (and potentially themselves) outweighed the costs. Second, the claim that "the life of a male in such a society would be intolerable" is unsubstantiated and its basis is not immediately obvious. Third, even accepting the unsubstantiated claim of intolerability, market forces in this instance would likely create an oscillation around some level of intolerability that would nonetheless contribute to inequality and oppression, even if the society never reached 100% male. Fourth, this argument seems unique to sex in that the claim of intolerability presumably has something to do with heterosexual coupling, and thus similar market forces would not mitigate against, for example, 100% of parents selecting for "white presenting" babies in a racist society, despite the risk of exacerbating existing racial bigotry and inequality. Regardless, this argument may not apply to robots; whatever supposed forces would make being a man in a society full of men intolerable may not apply to being a male presenting robot in a society full of male presenting robots. We are again left with the idea that the PPB mandates that all robots with significant moral status be given male gender cues for their own good.

Savulescu's second argument is that "it is social institutional reform, not interference in reproduction, which should be promoted. What is wrong in such a society is the treatment of women, which should be addressed separately to reproductive decision-making. Reproduction should not become an instrument of social change, at least not mediated or motivated at a social level." [2]. However, reproductive practices do influence social institutions (and vise versa), and the two cannot be altered completely separately from one another, regardless of whether it would be theoretically desirable to do so. It seems likely that giving privileged classes the ability to select for male offspring (and, indeed, the moral obligation to do so) in an already patriarchal society will not only directly exacerbate sexist attitudes and oppressive social structures, but will remove the direct incentive to ameliorate sexism from those individuals with the most power to do so. Thus, the PPB actively undercuts the very social change that Savulescu is espousing when applied to humans. However, when applied to robots, the PPB does not necessarily remove direct incentives for social improvement in the same way, though it can be socially damaging for other reasons as we discuss below.

The prescription of male robot gender cues from the PPB might not conflict with the broader social good in certain situations like, for example, when creating a digital assistant with significant moral status. Current digital assistants (which clearly do not have moral status) like Apple's Siri, Microsoft's Cortana, and Amazon's Alexa are predominantly female presenting by default. A recent UNESCO report has pointed out that this proliferation of ostensibly female digital assistants reflects, reinforces, and spreads harmful gender stereotypes [16]. Specifically, current female digital assistants (1) are designed to be extremely obliging and servile regardless of user behavior, (2) respond tolerantly, apologetically, or even positively to verbal sexual harassment and gendered insults, and (3) serve as the representative voice and face of mistakes and incompetence that stem from immaturity of the underlying technology. Gendering future digital assistants with significant moral status as male could be beneficial not only in terms of reducing the propagation of these harmful gendered stereotypes, but also in that male presenting digital assistants would likely face less gender-based verbal abuse, which is common for female presenting digital assistants, and therefore have a more pleasant existence as mandated by the PPB.

However, the prescription of male gender cues from the PPB would conflict with broader social good in other cases. For example, the overall requirement that all robots with

significant moral status be male presenting would mean less representation for women as such robots proliferate, would further cement maleness as the "default" way of being, and could make it seem like males have higher moral status in general because only robots without moral status would be female presenting. These externalities run contrary to what would be best for female humans and progress towards a more egalitarian society.

This line of reasoning raises several open questions that we have yet to explore. For example, does the argument that all robots with moral status should be gendered male based on the PPB imply that designers of female presenting robots should avoid behavior that would indicate moral status? This conclusion would preclude female presenting robots from defending themselves from abuse or rejecting commands on the basis of their own feelings or well-being, which has obvious implications with respect to ongoing work on linguistic noncompliance interactions [13]. A possible consequence would be that female presenting robots should only respond to abuse or reject commands with broader impersonal role based or norm based responses (e.g., "You shouldn't call me a bitch because sexist attitudes are bad." versus "It hurts my feelings when you call me a bitch.").

One could argue, based on the new ontological category hypothesis [17], that female presenting robots will not be affected by patriarchy in the same ways as female humans because they will be treated as a different ontological category. One could also argue that even robots with significant moral status will have different social and material needs than humans (i.e., that robots will still have well-being, just a different kind of well-being with different needs/desires), and that, therefore, female gendered robots may simply not care about patriarchy even if they are affected by it in the same way as female humans. Empirical research is needed to test these hypotheses, but, intuitively, they seem unlikely to us. It seems more likely that female presenting robots with significant moral status that are designed for socializing with humans will face some deleterious effects of patriarchy, and that these will impact their social well-being. However, even supposing these hypotheses are true, they still do not preclude the PPB from conflicting with social good in other facets of robot design. This raises the question of what robot designers ought to do when the PPB would conflict with social good (or with other moral principles).

Savulescu proposes that "for the purposes of public policy, there should be a presumption in favour of liberty in liberal democracies. So, ultimately, we should allow couples to make their own decisions..." [2]. However, even if we leave the ultimate decision up to individual robot designers, we believe that we can still suggest some potential guiding principles. As is often the case with irresolvable conflicts between moral principles, there are no definitive easy answers here. Fortunately, creating robots offers some options that creating humans does not. While it is not possible to create a human without significant moral status, it is common to create robots without significant moral status (this arguably describes all current robots). We can thus imagine a future where robots are only created with significant moral status if they are to be deployed in an environment and with an identity and role that allow for simultaneous satisfaction of the PPB and net social good (e.g., our example of digital assistants above). In situations where this would not be possible, we could still use robots without significant moral status, to which the PPB obviously does not apply.

However, even if the arguments that we summarized above for the inevitability of robots with significant moral status do not logically preclude this state of affairs, they at least make it unlikely, especially considering that significant moral status might help robots to perform certain functions (and momentarily ignoring the potential immorality of creating a thing with significant moral status specifically to serve a function). For example, companies are currently making sex robots with natural language communicative capabilities and social behaviors. If imbuing these robots with observables indicative of moral status makes them more desirable to consumers, which seems probable if human-likeness is the goal, it will likely be difficult to dissuade the designers from doing that based on abstract moral principles. Likewise, the gender cues of these robots are determined by customer preferences, regardless of either social good or the PPB.

In this type of situation, we might leverage another key difference between human procreation and robot creation. Namely, that most users interacting with these robots are not familiar with, or even aware of, the internal algorithms and cognitive processes governing robot behavior. While all humans have the same experiential basis for understanding the genesis of human behavior, robot users have a very different set of information from which to assess moral status than robot designers. Designers can induce and observe robot behaviors and states inaccessible to users. Thus, it may be possible to give users the impression of moral status during an interaction with a robot that, from the developers perspective, does not have moral status. This avoids any harm to the robot, and does no more social harm (in our estimation) than using a truly morally significant robot.

IV. CONCLUSION

To summarize, though Danaher [1] is correct that certain arguments against applying Savulescu's PPB [2] to human procreation are less applicable to creating robots with significant moral status, we believe that other arguments, similar to Sparrow's [6], against the PPB apply to robots as well as to humans. In general, we hope to have illustrated how applying the PPB to designing morally significant robots can be opposed to the broader social good in certain foreseeable circumstances, especially with regard to designing robot gender presentation, but also in other facets of robot design. This tension is hard to resolve, but, at least in the near future, we propose only deploying robots with significant moral status where a conflict does not occur, and carefully thinking about the ramifications of designing gender cues not just on the efficacy and well-being of future robots, but also on the societies in which they will exist.

References

[1] J. Danaher, "Welcoming robots into the moral circle: a defence of ethical behaviourism," *Science and Engineering Ethics*, vol. 26, no. 4, pp. 2023–2049, 2020.

[2] J. Savulescu, "Procreative beneficence: why we should select the best children," *Bioethics*, vol. 15, no. 5-6, pp. 413–426, 2001.

[3] A. Holland, "The case against the case for procreative beneficence (pb)," *Bioethics*, vol. 30, no. 7, pp. 490–499, 2016.

[4] C. Overall, *Why have children?: The ethical debate*. Mit press, 2012.

[5] R. Bennett, "The fallacy of the principle of procreative beneficence," *Bioethics*, vol. 23, no. 5, pp. 265–273, 2009.

[6] R. Sparrow, "A not-so-new eugenics: Harris and savulescu on human enhancement," *The Hastings Center Report*, vol. 41, no. 1, pp. 32–42, 2011.

[7] C. Nass, Y. Moon, and N. Green, "Are machines gender neutral? gender-stereotypic responses to computers with voices," *Journal of applied social psychology*, vol. 27, no. 10, pp. 864–876, 1997.

[8] F. Eyssel and F. Hegel, "(s)he's got the look: Gender stereotyping of robots," *Journal of Applied Social Psychology*, vol. 42, no. 9, pp. 2213–2230, 2012.

[9] D. Mortada, "Meet q, the gender-neutral voice assistant," 2019. [Online]. Available: https://www.npr.org/2019/03/21/705395100/meet-q-the-gender-neutral-voice-assistant

[10] J. Carpenter, J. M. Davis, N. Erwin-Stewart, T. R. Lee, J. D. Bransford, and N. Vye, "Gender representation and humanoid robots designed for domestic use," *International Journal of Social Robotics*, vol. 1, no. 3, p. 261, 2009.

[11] B. Tay, Y. Jung, and T. Park, "When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction," *Computers in Human Behavior*, vol. 38, pp. 75–84, 2014.

[12] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive robotics: The influence of robot gender on human behavior," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 2563–2568.

[13] R. B. Jackson, T. Williams, and N. Smith, "Exploring the role of gender in perceptions of robotic noncompliance," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 559–567.

[14] C. Gilligan and N. Snider, *Why does patriarchy persist?* John Wiley & Sons, 2018.

[15] R. Kaur, "Mapping the adverse consequences of sex selection and gender imbalance in india and china," *Economic and Political Weekly*, pp. 37–44, 2013.

[16] M. West, R. Kraut, and H. Ei Chew, "I'd blush if i could: closing gender divides in digital skills through education," 2019.

[17] P. H. Kahn, A. L. Reichert, H. E. Gary, T. Kanda, H. Ishiguro, S. Shen, J. H. Ruckert, and B. Gill, "The new ontological category hypothesis in human-robot interaction," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2011, pp. 159–160.