A BEHAVIORAL THEORY OF ROBOT RIGHTS

N. F. SUSSMAN*

ABSTRACT

What are the precise conditions under which we ought to ascribe fundamental rights to robots? This paper addresses the moral and legal status of artificially intelligent beings—a problem existing at the convergence of ethics, law, politics, and technological advancement—and suggests one potential solution that is both practice-oriented and supported by robust philosophical analysis. I begin by surveying the answers provided by prominent theorists working within and outside of the machine-ethics literature. The dominant propositions can broadly be categorized into what I call: (a) the "criterion of humanity," which holds that only human beings can possess legal rights in the political society we have constructed; and (b) the "criterion of moral agency," which holds instead that only moral agents can possess such rights. I find that each of these positions is untenable due to problems ranging from conceptual inconsistency to postulation that cannot be empirically verified. I then articulate and defend an alternative position, which I call the "criterion of behavioral symmetry." This position suggests that an intelligent machine ought to be granted fundamental rights if it becomes behaviorally indistinguishable from at least one human being, without any further requirements. I conclude that, although no machine may currently satisfy the criterion of behavioral symmetry, it seems plausible that a sufficiently developed robot could meet this requirement in the future, and it would be exceedingly difficult to justify the philosophical position that such a being should not have many of the same rights as humans.

I. INTRODUCTION

Advances in artificial intelligence technology have raised increasingly prominent questions about the ethical and legal relevance of machines. The developing philosophical literature that attempts to answer these questions

* J.D. 2020, University of Southern California Gould School of Law; M.Sc. Philosophy 2017, The London School of Economics and Political Science; B.A. Political Science 2016, University of Western Ontario. This essay was originally conceived as my master's dissertation at The London School of Economics and Political Science, and was submitted for that purpose in an earlier form during the fall of 2017. I thank my family and friends, and in particular my parents, for their support, guidance, and patience as I continuously requested their time and feedback to help develop the arguments in this paper. I also thank the faculty members in LSE's Department of Philosophy, Logic and Scientific Method, including Bryan W. Roberts, J. McKenzie Alexander, and Richard Bradley, for their wisdom and mentorship in the pursuit of crafting clear and convincing philosophical analyses. Finally, I thank the team of editors at the *Southern California Interdisciplinary Law Journal* for their considerate work throughout the publication process.

can generally be divided into two broad categories. The first category debates the values that ought to be considered and applied in the design of machines that may eventually be capable of making morally significant decisions. The second category includes normative and conceptual analyses of the moral and legal standing of the machines themselves to determine how they ought to be treated by people.[1] My argument in this Article focuses on the latter category, as I intend to address the philosophical issue of whether humanlike robots, or "androids," should be granted fundamental rights. Specifically, my analysis proposes and defends one practically oriented answer to the following question: What is at least one precise sufficient condition under which we ought to ascribe fundamental legal rights to intelligent androids?

The question of whether androids ought to be granted ethical or legal rights is hardly new. References to the possibility of robot rights are frequently made throughout both the vast literature of machine ethics and the even broader literature of philosophy, especially when acknowledging robot rights as a potential implication of ascribing some special moral status to intelligent machines.[2] The topic has also been subject to some analysis in legal scholarship, which has grappled not only with how androids could or should be cognized within existing legal frameworks, but also with how the frameworks would need to be amended in order to accommodate such beings.[3] Additionally, and perhaps most importantly, the topic has captured and retained the interest of the general public, as we can infer from the widespread popularity of films and television shows such as *Westworld,*[4] *Blade Runner: 2049*,[5] *Ex Machina*,[6] *Humans*,[7] and *Her*,[8] to name just a few. But the academic discourses on this matter remain highly disparate, and to the extent solutions are suggested, they often imply wildly different results.[9]

---

[1] DAVID J. GUNKEL, ROBOT RIGHTS 1–2 (2018) (containing a similar and helpful explication of this divide in the "machine ethics" literature).

[2] *See, e.g.*, Joshua C. Gellers, RIGHTS FOR ROBOTS (2021); Wolfgang M. Schröder, *Robots and Rights: Reviewing Recent Positions in Legal Philosophy and Ethics*, *in* ROBOTICS, AI, AND HUMANITY 191 (J. von Braun et al., eds., 2021); Jacob Turner, ROBOT RULES: REGULATING ARTIFICIAL INTELLIGENCE (2019); Herman T. Tavani, *Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights*, 9 INFORMATION 73 (2018); Mark Coeckelbergh, *Robot Rights? Towards a Social-Relational Justification of Moral Consideration*, 12 ETHICS & INFO. TECH. 209 (2010); David Levy, *The Ethical Treatment of Artificially Conscious Robots*, 1 INT. J. SOC. ROBOTICS 209 (2009); Peter M. Asaro, *What Should We Want from a Robot Ethic?* 6 INT'L. REV. INFO. ETHICS 9 (2006); Phil McNally & Sohail Inayatullah, *The Rights of Robots: Technology, Culture and Law in the 21st Century*, 20 FUTURES 119 (1988); *see also* Jamie Harris & Jacy Reese Anthis, *The Moral Consideration of Artificial Entities: A Literature Review*, 27 SCI. & ENG'G ETHICS 53 (2021) (engaging in an extensive and recent literature review of academic scholarship on the ethical or moral status of robots and broader AI systems).

[3] For example, could androids be "legal persons" even if not "natural persons," similar to corporations? *See, e.g.*, ALAIN BENSOUSSAN & JÉRÉMY BENSOUSSAN, DROIT DES ROBOTS [THE LAW OF ROBOTS] (2015); *see also* Belinda Bennett & Angela Daly, *Recognising Rights for Robots: Can We? Will We? Should We?*, 12 L., INNOVATION & TECH. 1 (2020).

[4] *Westworld* (HBO 2016).

[5] BLADE RUNNER: 2049 (Warner Bros. Pictures 2017).

[6] EX MACHINA (Universal Pictures 2015).

[7] *Humans* (AMC 2015).

[8] HER (Warner Bros. Pictures 2013).

[9] There are two illustrative examples with drastically divergent approaches. The first is encapsulated in the findings of Jennifer Robertson, which describe how Japanese society generally seems to welcome the idea that some notion of personhood could be attributed to robots (at least relative to developments in other post-industrial nations). *See, e.g.*, Jennifer Robertson, *Human Rights vs. Robot Rights: Forecasts from Japan*, 46 CRITICAL ASIAN STUD. 571 (2014). Indeed, Robertson reports on how the Japanese state actively encourages human-robot codependency—perhaps more than dependency between Japanese citizens and human foreign nationals—and how the robot 'Paro' was even granted its own *koseki* (a sort

Moreover, research indicates that after filtering out casual references and promissory comments, there is relatively little direct philosophical analysis affirmatively suggesting any practically actionable conditions under which we should feel compelled to grant rights to machines.[10] As machines become more sophisticated and androids become increasingly humanlike, this apparent gap in the now-speculative literature will eventually require filling, particularly given that public opinion studies suggest a desire for guidance.[11] My arguments in this Article are motivated by the growing need to expand the volume of philosophically robust theories offering practice-oriented proposals to address this issue.

The analysis presented here rests on the following axioms: (1) humans and androids are both categories of "entities," and (2) human entities possess fundamental legal "rights." My suggestion is that if androids become behaviorally indistinguishable from humans, society ought to grant fundamental legal rights to androids as well. This proposal is therefore situated within the emerging school in machine ethics sometimes referred to as "behaviorism." My argument takes the following logical form:

> **Premise I**: If an entity meets the criterion to be eligible for rights, that entity ought to be granted fundamental legal rights, regardless of the entity's other properties.
>
> **Premise II**: If an entity is behaviorally indistinguishable from at least one human, we ought to find that the entity meets the criterion to be eligible for rights.
>
> ---
>
> **Conclusion:** If an android is behaviorally indistinguishable from at least one human, the android ought to be granted fundamental legal rights, regardless of any extant differences between humans and androids.

Given that the overall argument is logically valid on its face, most of my analysis is devoted to normative and conceptual arguments supporting the soundness of its two premises.

---

of membership in Japan's household registry) which, although carrying no "legal force," connotes significant cultural meaning. *Id.* The second is the analysis provided by Günther, which advocates for a repurposing of ancient Roman law in order to essentially supply a legal framework that would recognize robots as "slaves." *See, e.g.*, JAN-PHILIPP GÜNTHER, ROBOTER UND RECHTLICHE VERANTWORTUNG: EINE UNTERSUCHUNG DER BENUTZER- UND HERSTELLERHAFTUNG. (2016).

[10] Even GUNKEL, *supra* note 1, at 10–30, who provides one of the most comprehensive analyses of the philosophical issues presented by the concept of "robot rights" to date, intentionally avoids questions regarding a "practical framework for devising and writing policy" and focuses instead on the more abstract "opportunities and challenges made available by the questions concerning robot rights" from a largely linguistic and critical-theoretic perspective.

[11] *See generally* Maarte M.A. de Graaf et al., *Who wants to grant robots rights?*, HRI '21 COMPANION: COMPANION OF THE 2021 ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION 38 (2021) (publishing survey results indicating that online human respondents varied in their attitudes toward the possibility of robot rights based on age and experience, as well as expectations about the potential cognitive abilities of robots); Gabriel Lima et al., *Collecting the Public Perception of AI and Robot Rights*, 4 PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION CSCW2 (2020) (publishing survey results indicating that online human respondents would intuitively prefer to deny the status of legal personhood to an AI system or robots for all purposes, *other* than the right to protection from cruelty, but that attitudes toward robot personhood generally became more positive when further information and argumentation was provided).

I begin by defining the central concepts invoked throughout this analysis. Section II defines "intelligent machines" and "androids," while Section III defines "fundamental rights" and explains the analytical difference between "having" legal rights and "being eligible" for such rights. In Section IV, I provide an argument to support Premise I of my overall thesis—any entity that meets the criterion to be eligible for fundamental rights ought to be afforded those rights under the law, regardless of the entity's other properties. Next, Sections V, VI, and VII each consider one possible criterion under which an entity may be eligible for rights. Sections V and VI present the criteria of "humanity" and "moral agency," respectively, and I argue that both should be dismissed due to various conceptual and practical weaknesses. My arguments include responses to several influential perspectives in machine ethics, including those of Batya Friedman and Peter Kahn,[12] Arthur Kuflik,[13] David Levy,[14] and Luciano Floridi and John Sanders, among others.[15] Section VII articulates and defends an alternative theory: "the criterion of behavioral symmetry." This criterion forms Premise II of my overall thesis: any entity that is behaviorally indistinguishable from a human ought to be considered eligible for fundamental rights. Finally, in Section VIII, I briefly consider the factual question of whether an android could actually meet the criterion of behavioral symmetry.

Before continuing, I should note that many of the existing arguments in machine ethics are somewhat disparate. The philosophical debates in this area often use vastly different concepts, assumptions, and methodological tools, such that many influential claims are not easily comparable at first glance. Another way of expressing this complication is to say that, at present, no single paradigm governs our inquiry into machine ethics and robot rights. For this reason, I present many of the existing perspectives by rationally reconstructing them such that they can be placed into a coherent conversation.

II. ARTIFICIAL INTELLIGENCE AND ANDROIDS

I define artificial intelligence ("AI") as the process of "thinking" by which input data are computed and stored to produce functional outputs, where the process is ultimately executed by means of human-designed programs, and where such programs consist in formal systems that manipulate information to produce outputs with a level of sophistication that may equal or surpass ordinary human computational capabilities. I intend for the working definition of AI described here, and explained in greater detail below, to be relatively uncontroversial. Indeed, this definition is consistent with many influential analyses found throughout the AI ethics literature.[16]

---

[12] Batya Friedman & Peter H. Kahn, Jr., *Human Agency and Responsible Computing: Implications for Computer System Design*, 17 J. SYS. SOFTWARE 7 (1992).

[13] Arthur Kuflik, *Computers in Control: Rational Transfer of Authority or Irresponsible Abdication of Autonomy?*, 1 ETHICS & INFO. TECH. 173 (1999).

[14] Levy, *supra* note 2.

[15] Luciano Floridi & John Sanders, *On the Morality of Artificial Agents*, 14 MINDS & MACHS. 349 (2004).

[16] *See, e.g.*, Patrick Chisan Hew, *Artificial Moral Agents Are Infeasible with Foreseeable Technologies*, 16 ETHICS & INFO. TECH. 197 (2014); Frances S. Grodzinsky, Keith W. Miller & Marty J. Wolf, *The Ethics of Designing Artificial Agents*, 10 ETHICS & INFO. TECH. 115 (2008); Andreas Matthias, *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, 6 ETHICS & INFO.

As a basic example, the simple program depicted in Appendix A is written in the coding language Python 3; this program provides the instructions for a computational machine capable of reading such script to calculate the value of pi (π) using the method originally developed by Gottfried Wilhelm von Leibniz. This formal, human-designed program would enable a capable machine to perform the calculation with equal, if not far greater, speed and accuracy when compared to that of the average human. This sort of sophistication distinguishes what I conceived of as AI programs from simpler machine-implemented programs, such as traditional air conditioners and toasters.

Moreover, in recognition of the most substantial recent advances in computational technology, I should specifically note that my working definition of AI would include systems that execute via symbolic machine learning algorithms or "connectionist" neural nets, which operate by means of programs that self-update according to the external "feedback" such programs may receive on their prior outputs.

I define "intelligent machines" or "robots" as the individuated, human-designed agents that function by means of AI programs. I then define "androids" as intelligent machines that are designed to be human-like in both appearance and overall functioning. I do not include machines whose programs may eventually operate by means of "whole-brain emulation" within my definition of androids. This is because I am concerned that potential AI substrates of that kind may be particularly under-researched at this time and that they would introduce additional theoretical challenges beyond the scope of the present analysis. I have chosen to focus on androids rather than all intelligent machines more broadly because, at least intuitively, robots with human-like functionality seem to elicit the most perplexing ethical—and perhaps, by extension, legal and political—questions.

III. FUNDAMENTAL RIGHTS

The "fundamental rights" with which I am concerned for purposes of this analysis are legal rights. That is, as I discuss the concept of "rights" and "having rights" throughout this paper, I have in mind the sort of rights that are created and enshrined in constitutions, treaties, statutes, and institutional practices, and that are tangibly enforced in courts or other tribunals. My definition of fundamental rights does not include the more abstract set of moral or ethical rights that, while relevant to the problems of ethical decision-making, lack dispositive authority in legal disputes between persons. The more abstract set of moral and ethical rights remain important to this analysis because they help us to determine whether an entity is "eligible" for fundamental rights; but these more abstract rights tell us

TECH. 175 (2004). Symbolic AI systems operate algorithmically: sets of rules manipulate symbols, which correspond to data as defined in a table. Machine learning algorithms expand on this process by allowing for the tables and rules to self-modify according to experiential feedback. In the case of neural nets, the system's updating is conducted by adjusting the weighted connections between sets of "neuron" analogs. Information that is manipulated by neural nets is not represented by discrete symbols; rather, this information is spread across the entire net. Matthias, *supra*.

nothing about whether an entity actually has such rights. I explore this distinction further below.

A. THE NATURE OF FUNDAMENTAL RIGHTS

For purposes of this paper, I understand fundamental rights as a socially constructed phenomenon. I do not follow natural rights theorists such as Locke[17] or Nozick[18] in theorizing that such rights have an independent metaphysical existence prior to their institutional development within societies. Instead, I follow the legal positivist tradition for these purposes and hold that fundamental rights exist because groups of people, either explicitly or implicitly, agree about and uphold the existence of fundamental rights.

The benefit of understanding fundamental rights as a social construction is that, when asking whether an entity has or ought to have rights, we are not required to locate some "mind-independent," "universal," or "objectively true" answer to the question. Rather, because rights are the product of human minds and institutions, we can understand truths about rights as facts imposed by intersubjective conventions, which obtain their normative force *as a result of* our intersubjective agreement. It follows that our conventions regarding which entities have or ought to have rights can change over time without damaging the conceptual or normative foundations of fundamental rights, provided there is adequate intersubjective agreement between people about these changes.

B. THE JUSTIFICATION FOR FUNDAMENTAL RIGHTS

Theorists sometimes disagree about the underlying justifications for the existence of fundamental rights. While the intricacies of this disagreement have been the subject of many classic arguments throughout the history of ethical, political, and legal theory, they are beyond the scope of this Article. For present purposes, I adopt the position that fundamental rights are justified simply because we have a widely shared moral intuition that any well-functioning liberal society, perhaps by definition, is required to uphold such rights for their own sake. This position is consistent with the Kantian tradition, which holds that rights are justified as ends in themselves rather than because they promote some other, extrinsic goal such as utility or wealth. This theory is sometimes called a "status-based," as opposed to "instrumental," theory about the justification of rights.[19] In sum, then, my philosophical position for purposes of this Article is that basic legal rights, such as those of liberty and self-determination, exist *because of social institutions*, and that the existence of basic legal rights is normatively *justified* due to our broad intersubjective agreement that there is a class of beings who hold a uniquely valued moral status that *inherently* warrants certain legal protections.[20]

---

[17] *See* JOHN LOCKE, THE SECOND TREATISE ON CIVIL GOVERNMENT (1689).

[18] *See* ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA (1974).

[19] Leif Wenar, *Rights*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2020), https://plato.stanford.edu/entries/rights/ [https://perma.cc/N68F-4TGY].

[20] This understanding of rights may be grounded in the meta-ethical doctrine of "constructivism," wherein moral obligations emerge from practical reasoning between individuals, and do not have any 'mind-independent' existence. JOHN RAWLS, A THEORY OF JUSTICE (1971); John Rawls, *Kantian*

C. WHAT DO FUNDAMENTAL RIGHTS INCLUDE?

To maintain a manageable scope for the present analysis, I limit my definition of "fundamental rights" so that it includes only the most straightforward and widely accepted set of basic liberties and legal protections currently afforded to individuals in liberal societies throughout the world. Drawing on a survey of several influential declarations of human and civil rights, I consider these fundamental rights to include liberty over the body and mind, self-determination, free expression, free movement, free association, and peaceful assembly. These are frequently thought of as at least some of the basic "negative liberties" held by individuals relative to others, including governments.[21] A sample of the primary documentary sources consulted includes the *United Nations Universal Declaration of Human Rights*,[22] the *United Nations International Covenant on Civil and Political Rights*,[23] the *United States Constitution*,[24] the *Charter of Fundamental Rights of the European Union*,[25] the *Grundgesetz* (German Constitution),[26] the *Xianggang Jiben Fa* (Hong Kong Constitution),[27] the *Constitution of India*,[28] and the *Constitution of the Republic of South Africa*.[29]

With this definition in mind, we can stipulate that non-human animals have generally *not* been granted fundamental rights, even though certain species of non-human animals are granted special legal protections under legislation such as the United States' *Preventing Animal Cruelty and Torture Act*[30] and the United Kingdom's *Animals (Scientific Procedures) Act*.[31] These protections for certain non-human animals do not rise to the level of fundamental rights under my definition of the term because these protections do not allow for basic liberties such as self-determination and free expression. We can also state somewhat non-controversially that non-animal entities, including machines of any kind, lack fundamental rights under current law. Although there can of course be "non-natural persons" who hold rights within our existing legal framework (such persons could typically include business organizations, trusts, and other associations or groups), these persons are generally not considered to have fundamental rights such as "liberty over the body and mind" in the sense that I have conceived of such concepts here.

Constructivism in Moral Theory, 77 J. PHIL. 515 (1980). This doctrine enables us to view rights as a socially constructed, although inherently morally justified, convention of liberal societies.

[21] *See* ISAIAH BERLIN, TWO CONCEPTS OF LIBERTY (1958).

[22] G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948).

[23] G.A. Res. 2200 (XXI) A, International Covenant on Civil and Political Rights (Dec. 16, 1966).

[24] U.S. CONST.

[25] Charter of Fundamental Rights of the European Union, 2012 O.J. C 326/391.

[26] Grundgesetz [GG] [Basic Law], translation at https://www.gesetze-im-internet.de [https://perma.cc/E5FZ-JMYJ].

[27] XIANGGANG JIBEN FA.

[28] India Const.

[29] S. AFR. CONST., 1996. Although there are certain additional rights common to these documents, I have included only those that I understand to be least controversial to avoid deviating beyond the scope of my argument in this Article.

[30] Preventing Animal Cruelty and Torture Act, 18 U.S.C. § 48.

[31] Animals (Scientific Procedures) Act, 1986, c. 14 (Eng.).

D. "HAVING RIGHTS" VERSUS "BEING ELIGIBLE FOR RIGHTS"

My conception of fundamental rights entails that the property of "having" fundamental rights is distinct from the property of "being eligible for" such rights. When I ask whether an entity *has* fundamental rights, I am asking whether it has fundamental *legal* rights; we must answer this question using strictly positivistic evidence—we need only consult a society's legal and political rules, practices, and institutions. To *grant* or *ascribe* fundamental rights to a new entity is to amend our rules, practices, and institutions such that the entity's fundamental rights have meaning and are protected under the law. Using such an analysis, we can state the empirical fact that humans have fundamental rights, while non-human entities, such as machines and non-human animals, do not.

But while the property of "having" fundamental rights is an empirical status, the property of "being eligible" for those rights is a moral one. An entity is considered "eligible" for rights if and only if the entity itself meets a specific criterion, or set of criteria, as formulated by the moral conventions of a society. It is therefore logically possible for an entity to be eligible for fundamental rights according to morality without actually *having* those rights according to law.

As an illustration, philosophers such as Tom Regan claim that some non-human animal species meet the criterion to be morally eligible for rights.[32] But although such theorists believe certain animal species are eligible for fundamental rights, such theorists clearly do not believe that those species actually have such rights as a result of their moral status. Indeed, the very goal of these theorists' arguments is to address the fact that those species presently lack fundamental rights as a matter of empirical observation.

I have been careful to differentiate the property of "having" fundamental rights from the property of "being eligible" for such rights because the distinction is central to the analysis that follows. In Section IV, I argue that any entity that meets the criterion of eligibility for rights ought to be granted those rights, regardless of its other properties. For clarity and to avoid unnecessary repetition, I will often refer to the class of entities that are eligible for rights as "set *E*" or simply "*E*." Accordingly, I will at times refer to the class of entities that actually have rights as "set *R*" or "*R*."

E. THE RESULTING CHALLENGE

When is an entity eligible for fundamental rights? That is, what *features* must an entity exhibit to be eligible? Because being eligible is a moral status, the *criterion* we should adopt for assessing an entity's eligibility is determined by moral reasoning. But just as there can always be disagreement about the requirements of morality, there can be competing views about *which* criteria embody the most appropriate conditions under which an entity ought to be considered eligible for fundamental rights.[33] Any argument about an entity's eligibility will be totally unpersuasive to those who measure eligibility according to a different criterion. It is therefore crucial that before debating whether androids are eligible for fundamental rights, we adopt at

[32] *See* TOM REGAN, THE CASE FOR ANIMAL RIGHTS (1983).
[33] Wenar, *supra* note 19; Coeckelbergh, *supra* note 2.

least one independently sufficient and widely acceptable criterion for eligibility. Identifying this criterion requires rigorous normative argumentation. I articulate and defend "the criterion of behavioral symmetry" in Section VII after first considering alternatives in Sections V and VI.

Additionally, identifying an appropriate criterion for eligibility is not, by itself, sufficient to form a complete argument that an entity ought to have fundamental rights. Rather, settling on such a criterion for eligibility is only one important step in the project. Because we have distinguished between being eligible for rights and actually having them, this distinction entails that any grand claim that some entity "ought to have rights" actually consists of three assertions. The first is a normative argument that any entity that is eligible for rights ought to be *granted* such rights under the law, regardless of any of the entity's other properties. The second is another normative argument that a certain set of features, from a moral perspective, should comprise a sufficient condition to be eligible for rights—this argument is the subject of Sections V, VI, and VII. The third is a descriptive assertion that the entity in question meets the agreed-upon sufficient condition for eligibility, which is considered in Section VIII. I undertake to defend the first of these assertions in the following Section.

IV. PREMISE I: BEING CONSISTENT IN ASCRIBING RIGHTS

I have adopted the "status-based" philosophical position that fundamental rights are justified because of the moral status of the beings they protect. Specifically, these legal rights ought to be upheld because they protect entities that hold the moral status of "being eligible for rights." I do not suggest that there is anything "universal," "objective," or "natural" about this moral status. Rather, I conceive of the conditions for eligibility as belonging to socially constructed reality. As such, the moral status of being eligible for rights is upheld as a matter of convention, and the criterion, or set of criteria, for eligibility can change with sufficient intersubjective agreement.

Given that the existence of rights is justified by the existence of entities that we deem morally eligible for rights, it must follow that *any* entity that meets the conditions for eligibility ought to be granted those rights, regardless of the entity's other properties. If we only actually grant membership in R (the class of entities that actually have rights) to some of the members of E (the class of entities that are eligible for rights), this dynamic amounts to unjustified arbitrariness or even randomness in ascribing rights to certain lucky beings. This arbitrariness would be analogous to preparing a very large meal on the grounds that everyone in a group is hungry, but subsequently deciding to serve food to only a few group members—the justification for the action is undermined by its inconsistent execution. From this perspective, a failure to be consistent in ascribing rights entails that their existence is not truly justified at all, and this state of affairs seems contrary to our principles of justice and our intuitive notion of fundamental rights as a consistent, justified, and fair institution characteristic of well-founded liberal societies.

Our inherent liberal value for being justified and consistent in ascribing fundamental rights can be elaborated by way of examples. It is a fact that throughout a significant portion of American history, many minority groups have not been granted fundamental rights (that is, have not been members of *R*) in the United States. The historical exclusion of African-Americans, at least from the complete set of fundamental legal rights contemplated in Section III, is paradigmatic in this respect. It is also a fact that, over the course of the nineteenth, twentieth, and twenty-first centuries, many of these same groups have gradually become members of *R* through hard-fought legislative and social reforms (as many would acknowledge, with significant legal and social progress still to be made). If we did not care about being consistent and justified in ascribing rights, we would coldly observe these exclusions and subsequent reforms as a factual curiosity of history. Instead, most of us share a pervasive sense that these facts represent an incredibly important historical injustice, one worthy of intense criticism and which we must strive to avoid repeating at all costs. This sense of injustice arises from our judgment that, even before the persons in these groups were *granted* legal rights, they satisfied the moral criterion of *eligibility* for rights. Since they were members of *E*, but not *R*, society was *inconsistent* in its ascription of legal rights, thereby undermining the very justification for such rights in the first place.

From these beliefs, values, and intuitions about liberal justice, we can derive the normative principle that whenever an entity, such as a human or an android, is a member of *E*, but not *R*, this constitutes an injustice that ought to be rectified by adding the entity to *R* through institutional reforms. This principle applies for any member of *E*, irrespective of any other attributes that such a member may happen to possess. I expect this normative principle to be relatively uncontroversial, and it embodies Premise I of my overall argument.[34]

The greater complication is in identifying the most appropriate necessary and sufficient conditions for membership in *E*, and subsequently determining which entities do and do not meet the criterion that these conditions embody. These normative and empirical questions are the subject of the following Sections and occupy the majority of my analysis. I have thus far argued that if androids meet the criterion for belonging to *E*, they ought to belong to *R*, regardless of other unique properties that androids might possess. I now consider what exactly ought to be the criterion for belonging to *E*.

[34] F. Patrick Hubbard followed a similar line of reasoning when he asserted that "a denial of personhood to an entity with at least an equal capacity for personhood would be inconsistent and contrary to the egalitarian aspect of liberalism." F. Patrick Hubbard, *"Do Androids Dream?": Personhood and Intelligent Artifacts*, 83 TEMP. L. REV. 405, 417 (2011). Additionally, the premise presented here arrives at a similar conclusion to that described by Erica Neely when she discusses the issue of machine rights. Erica L. Neely, *Machines and the Moral Community*, 27 PHIL. & TECH. 97 (2014). Neely argues that entities are "moral patients" if they possess "interests," *id.* at 101, and points out that our "failure to acknowledge the moral standing of machines does not imply that they actually lack moral standing; we are simply being unjust in such cases, as we have frequently been before," *id.* at 106. Neely then proceeds to defend an attitude of "moral generosity" in determining whether to ascribe rights to new kinds of entities, given that greater social injustice will arise from an error in *failing* to grant such rights than an error in actually granting them. *Id.* at 109. *But see* Joanna J. Bryson, *Robots Should Be Slaves*, *in* CLOSE ENGAGEMENTS WITH ARTIFICIAL COMPANIONS 63–74 (Yorick Wilks ed., 2010) (arguing against adopting "moral generosity" toward machine entities given the potential individual and institutional costs to humanity, among other considerations).

V. THE CRITERION OF HUMANITY

Throughout this Section, and the two Sections that follow, I attempt to identify the criterion that ought to determine whether entities are eligible for fundamental rights. After considering and discarding the criteria of "humanity" and "moral agency," I claim that being "behaviorally indistinguishable from a human" is an acceptable sufficient condition for eligibility. I will devote substantial effort to discrediting the criteria of humanity and moral agency—and subsequently supporting the criterion of behavioral symmetry—because the latter supplies Premise II of my overall argument.

Confronted with the challenge of identifying the most appropriate criterion for membership in *E*, we might be tempted to adopt the following position:

> **The criterion of humanity:** An entity is eligible for rights if and only if the entity is a human being.

This seems to be the implied position of many theological philosophers, such as Thomas Aquinas,[35] who hold human beings to be morally unique among God's creations. This criterion is also compatible with the fact that humans are the only entities that are currently granted fundamental rights. If accepted, this criterion would analytically imply that non-human entities could not be eligible for rights, thereby denying any claims to rights for machines.

The problem with the criterion of humanity is that it does not fit well with many of the modern political and academic discourses concerning the topic of fundamental rights. From these discourses, it seems clear that we do not consider humans eligible for rights simply *because* we are human. Rather, we tend to justify humanity's eligibility for rights on the basis of certain underlying properties that humans possess. A sample of these properties would include our "inherent" rationality, our capacity to feel pleasure and pain, and our experience of self-consciousness.[36] Consider, for instance, Article I of the *Universal Declaration of Human Rights*, which justifies "free and equal" rights on the basis of our "reason and conscience."[37] On the assumption that these underlying properties, rather than simply the property of "being human," constitute the proper necessary and sufficient conditions for membership in *E*, we have developed vast literatures discussing the possibility of non-human rights that include robot rights, animal rights, group rights, and rights for ecosystems.[38] In other words, it proves difficult to find evidence that people in general actually find the criterion of humanity to be appealing or persuasive, and it proves relatively easy to find evidence to the contrary.

Once the criterion of humanity is discarded as undesirable, we are left with the question of exactly *which* human properties ought to constitute the

---

[35] THOMAS AQUINAS, *Summa Theologica* (c. 1274), *reprinted in* BASIC WRITINGS OF ST. THOMAS AQUINAS: VOLUME 1 (Anton C. Pegis ed., 1997).

[36] Coeckelbergh, *supra* note 2.

[37] G.A. Res. 217 (III) A, Universal Declaration of Human Rights, art. I.

[38] Coeckelbergh, *supra* note 2; REGAN, *supra* note 32; *see also* PETER A. FRENCH, COLLECTIVE AND CORPORATE RESPONSIBILITY (1984).

necessary and sufficient conditions for membership in *E*. Although humans are clearly judged to be members of *E* and rocks are clearly not, there is a vast number of ways in which humans differ from rocks. Which of these differences ought to be considered salient for determining each entity's moral status and thus their eligibility for rights?

To begin addressing this question, Nick Bostrom and Eliezer Yudkowsky point out that biological factors are not normally considered relevant in assessing an entity's moral status.[39] They encode this convention into two principles. The first is the "principle of substrate non-discrimination," which holds that "if two beings . . . differ only in the substrate of their implementation, then they have the same moral status."[40] The second is the "principle of ontogeny non-discrimination," which holds that "if two beings . . . differ only in how they came into existence, then they have the same moral status."[41] These principles imply that neither the distinctive human biology in its current form, nor the evolution of *homo sapiens* over millions of years by the process of natural selection, constitute relevant factors in our judgment that humans, but not rocks, are eligible for rights.[42] And these principles seem to provide appropriate starting points, or "guard rails," for purposes of identifying the most salient qualities when assessing an entity's moral status, looking past simply "being human." The alternative would be to accept that we should ascribe fundamental legal rights to entities based only upon *how* they exist or came into existence—two sets of scientific qualities that almost no one seems to cite when justifying the advent of rights—and qualities that, if anything, could be and have been used to formulate irrelevant, unhelpful, and rhetorically dangerous differentiating factors even among humans.

VI. THE CRITERION OF MORAL AGENCY

Western intellectual history is permeated with the notion that human beings hold the highest moral status because we possess the property of "moral agency." On this assumption, it is exceedingly common to assume that "moral agency" ought to constitute the necessary and sufficient condition for membership in *E*. This position can be formulated as such:

> **The (naïve) criterion of moral agency:** An entity is eligible for rights if and only if the entity is a moral agent.

---

[39] *See* Nick Bostrom & Eliezer Yudkowsky, *The Ethics of Artificial Intelligence*, *in* THE CAMBRIDGE HANDBOOK OF ARTIFICIAL INTELLIGENCE 316 (2014).

[40] *Id.* at 322–23.

[41] *Id.* at 323. It is worth noting that both of Bostrom and Yudkowsky's principles are subject to some debate. For example, Lantz Miller has argued that "maximally humanlike automata" must be ontologically distinct from human beings in at least one sense, even if similar in almost every other respect. Lantz Fleming Miller, *Granting Automata Human Rights: Challenge to a Basis of Full-Rights Privilege*, 16 HUM. RTS. REV. 369 (2015). The distinction is that while human beings evolved through a process that is "normatively neutral," all machines come into existence from the purposive, norm-driven efforts of other beings. *Id.* Miller argues that this difference is sufficient to justify denying rights even to "maximally humanlike automata." *Id.*

[42] That is, these properties do not constitute relevant factors *in and of themselves*. If a property such as "consciousness" is a salient difference, and consciousness is the result of human biology, then the biological factor is only indirectly relevant.

Let us define "moral agency" according to what Kenneth Himma calls the "standard view."[43] An "agent" is an entity capable of performing purposeful actions within its environment. These actions can involve other agents as well as non-agentive objects. A "moral agent" may then be defined as an agent that is appropriately held morally praiseworthy or blameworthy by others for its actions. Henceforth, the capability for being held morally praiseworthy or blameworthy is summarized as being "morally responsible."[44] Since an entity can only be a moral agent if it is capable of being held morally responsible, we can restate the naïve criterion of moral agency as follows:

> **The (restated) (naïve) criterion of moral agency:** An entity is eligible for rights if and only if the entity is an agent that is held morally responsible for its actions.

It seems immediately clear that the naïve criterion of moral agency does not plausibly articulate a *necessary* condition for membership in *E*. After all, we normally consider infant children and vegetative or mentally incapacitated persons to have the moral status for membership in *E*, although we may not intuitively hold these individuals to be morally responsible. Facing these exceptions, it seems we can exercise a degree of charity by revising the naïve criterion of moral agency so that it "saves the phenomena" by including every entity that we already consider to be a member of *E*. Common to infants, as well as vegetative or mentally incapacitated humans, is a teleological relation to moral agency—in each case these beings will be, may be, could have been, or were at some time moral agents. We can thus expand the criterion in the following way:

> **The (revised) criterion of moral agency:** An entity is eligible for rights if and only if the entity is, was, will be, may be, or could have been a moral agent.

This revised criterion (henceforth "the criterion of moral agency") seems at first to work as a necessary *and* sufficient condition for membership in *E*. Intuitively, it proves difficult to imagine an entity that would be eligible for rights without meeting this criterion, while it proves equally difficult to imagine an entity that meets this criterion without being eligible for rights.

The criterion of moral agency is also widely accepted as a reason for why androids cannot be considered eligible for rights. While there are generally no objections to calling androids agents, it is highly controversial to attribute moral responsibility to a machine. And by arguing that intelligent machines cannot be held morally responsible for their actions, many theorists claim that androids cannot satisfy the criterion of moral agency.[45]

Throughout the remainder of this Section, I argue that the criterion of moral agency suffers from serious deficiencies once we attempt to define its concepts precisely and that these deficiencies ought to disqualify it as the

---

[43] Kenneth Einar Himma, *Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to be a Moral Agent?*, 11 ETHICS & INFO. TECH. 19 (2009).

[44] Andrew Eshleman, *Moral Responsibility*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2014).

[45] Hew, *supra* note 16; Steve Torrance, *Ethics and Consciousness in Artificial Agents*, 22 AI & SOC'Y 495 (2008); Grodzinsky et al., *supra* note 16; Kuflik, *supra* note 13; Friedman & Kahn, *supra* note 12.

ultimate criterion for membership in *E*. If we cannot formulate precise conceptions of moral agency and moral responsibility, we cannot reach definite conclusions about whether or not an entity is a moral agent. If we cannot definitively determine whether entities are moral agents, then the criterion of moral agency holds that no entity can ever be definitively considered eligible for rights. As this is an undesirable outcome, I propose that we replace the criterion of moral agency with "the criterion of behavioral symmetry" in Section VII.

Before defending my proposed criterion, I will devote some effort to exposing the weaknesses of the widely accepted criterion of moral agency. I consider three problematic attempts at precisely defining "moral agency" and "moral responsibility" in turn.

A. DEFINITION A: VOLUNTARY ACTION

Aristotle contended that agents could be held morally responsible for their actions if and only if those actions were *voluntary*, [46] and Andrew Eshleman's analysis suggests that this definition remains dominant among theorists of moral responsibility.[47] Indeed, defining moral agency in terms of voluntary action seems suitable because this definition fits comfortably with our existing intuitions about when someone can and cannot be held morally responsible. To clarify this intuition, consider the following thought experiment.[48] Imagine that Christina stands below a cliff. In one case, a boulder falls from the cliff due to the natural force of a storm and subsequently kills Christina. In another case, Henry, who is standing on the cliff next to the boulder, voluntarily chooses to push the boulder from the cliff and subsequently kills Christina. We tend to feel that Henry, but not the storm, is morally responsible for Christina's death. Aristotelian moral responsibility holds that we distinguish between the two cases because Henry acted voluntarily, whereas the storm did not.

Aristotle's identification of moral responsibility with voluntary action provides a premise from which several philosophers argue that machines cannot be moral agents.[49] These arguments posit that any machine operating by means of a formal program has no true choice about its functional responses to environmental stimuli—every conceivable output is determined, in some regard, by the machine's program. In these cases, it is claimed that the human programmer retains all moral responsibility for the actions of the machine.

Patrick Hew summarizes this position by describing machines as systems governed entirely by hierarchical sets of rules.[50] These sets of rules determine the system's behavior and are nested such that there are progressively higher-level rules that determine which lower-level rules obtain in the case of conflict. For any such system that is finite, it follows that there will be an ultimate meta-rule that governs the entire scope of the system's rule-based behavior. This ultimate rule cannot itself be subject to

[46] ARISTOTLE, THE NICHOMACHEAN ETHICS (1985).
[47] Eshleman, *supra* note 44.
[48] This thought experiment is adapted from the scenario posed by Friedman & Kahn, *supra* note 12.
[49] Hew, *supra* note 16; Grodzinsky et al., *supra* note 16; Kuflik, *supra* note 13.
[50] Hew, *supra* note 16.

any lower-level rules, and therefore cannot be created or modified by any aspect of the system's behavior; this rule must be created and modified by an external human programmer. This is even the case for machines that can self-modify their programming, as the process of modification is itself governed by meta-programs. So, the Aristotelian would view the agent who conceived of the meta-program—that is, the ultimate human programmer—rather than the entity implementing the program and its sub-programs, as the moral agent in these scenarios.

However, despite the popularity of Aristotle's theory, there are philosophical complications that arise from defining moral responsibility in terms of voluntary action. Specifically, Aristotle's understanding of voluntary action holds an agent's actions to be voluntary only if the agent possesses "free will" and therefore could have chosen to act differently. In other words, a "voluntary action" is an action that is ultimately originated by an agent's own choices without being coerced by any influences operating outside of the agent's cognitive field.[51] It follows that whenever an agent's actions are *not* voluntary in this sense, Aristotle would not hold the agent morally responsible. It follows further that, if free will did not exist at all, actions would *never* be voluntary, and there would be no Aristotelian moral agents.

The problem for Aristotelian moral responsibility, according to "incompatibilist" theorists, is that the existence of free will is not viable if we adhere to the modern worldview of "scientific determinism." Scientific determinism combines metaphysical commitments with a methodological scheme for explanation and holds that every entity exists, and every event necessarily occurs, as a result of antecedent causal conditions. These antecedent conditions are understood to be both the states of all matter and energy in the universe at any moment and a set of physical laws that govern how these states can change.[52] From the perspective of scientific determinism, our deliberations, choices, and actions are events that are embedded in a determinate causal chain that stretches throughout time. Accordingly, this perspective renders our notions of free will and voluntary action as folk psychological accounts of phenomena that are actually caused by complex, predetermined events in and outside of our brains. Therefore, incompatibilists hold that if we adopt scientific determinism, there can be no Aristotelian moral responsibility.

One might attempt to salvage Aristotelian moral responsibility by rejecting the metaphysics of determinism. This requires a position known as "metaphysical libertarianism" in which free will is claimed to exist and voluntary action is therefore possible.[53] The difficulty with this position, however, is that it suffers from several compelling counterarguments. For example, it seems that if we reject the claim that our choices are determined by prior conditions, we must be committed to the view that our choices are *indeterminate* or *random*.[54] Adopting this viewpoint over determinism may not therefore be helpful to Aristotelian moral responsibility since we do not

[51] *See* Timothy O'Connor, *Free Will*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2018).
[52] Eshleman, *supra* note 44.
[53] O'Connor, *supra* note 51.
[54] *See generally* A.J. AYER, PHILOSOPHICAL ESSAYS (1954).

traditionally conceive of voluntary actions as random any more than we conceive of them as determined by antecedent conditions. It seems inappropriate to call an action "voluntary" if it is understood to be the result of mere chance since the agent does not have sufficient *control* over the outcome. This counterargument concludes that, regardless of whether we adopt metaphysical libertarianism or determinism, we reach the same conclusion: no physical entity is capable of voluntary action.

Furthermore, it seems that, to a great extent, empirical evidence supports the position that human decision-making is indeed determined by factors operating outside of our conscious deliberations. Joshua Greene and Jonathan Cohen have argued that studies in the field of cognitive neuroscience increasingly support the conclusion that "mental" events, such as personal choice-making, can be identified with, and are caused by, physical events in the brain and can therefore be understood as part of a causal system extending beyond the individual mind.[55] Benjamin Libet et al. are largely credited with inaugurating this body of evidence,[56] although the implications of those particular experiments are avidly debated.[57] More recently, a study by M.C. Brower and B.H. Price posited a direct causal link between frontal lobe dysfunction in the brain and the frequency with which a person will "choose" to engage in violent behavior.[58] This suggests that the decision to act violently may be the result of prior determinate events in the brain, rather than "free" mental deliberation on the part of the agent. Citing this sort of evidence, Greene and Cohen claim that metaphysical libertarianism and Aristotelian moral responsibility are "threatened . . . pointedly."[59]

If we find the position of determinism more compelling than metaphysical libertarianism, or if we adopt certain deterministic interpretations of the results from fields such as neuroscience, we may be forced to discard the idea that our actions are voluntary in any traditional sense. This is the incompatibilist perspective. Although I do not intend to take any particular stance on these debates, I have presented the arguments in order to clarify the philosophical challenges involved with the concept of voluntary action. The ferocity of this debate creates serious problems for the acceptability of the Aristotelian definition of moral responsibility and for any argument that attempts to use the Aristotelian definition as a standard for assessing an entity's moral agency. Indeed, if the capacity for voluntary action is adopted as the standard that must be met for an entity to have moral agency—as Patrick Hew, Arthur Kuflik, Andreas Matthias, and others assume[60]—then there is no general agreement that humans, machines, or anything else in the physical universe can straightforwardly be considered a moral agent.

---

[55] Joshua Greene & Jonathan Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, 359 PHIL. TRANSACTIONS ROYAL SOC'Y B: BIOLOGICAL SCIS. 1775–85 (2004).

[56] Benjamin Libet, Curtis A. Gleason, Elwood W. Wright & Dennis K. Pearl, *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)*, 106 BRAIN 623 (1983).

[57] *See* ALFRED R. MELE, EFFECTIVE INTENTIONS: THE POWER OF CONSCIOUS WILL (2009).

[58] M.C. Brower & B.H. Price, *Neuropsychiatry of Frontal Lobe Dysfunction in Violent and Criminal Behavior: A Critical Review*, 71 J. NEUROLOGY, NEUROSURGERY & PSYCHIATRY 720 (2001).

[59] Greene & Cohen, *supra* note 55.

[60] *See* Hew, *supra* note 16; Kuflik, *supra* note 13; Matthias, *supra* note 16.

I anticipate the objection that, even if we dismiss the Aristotelian definition of moral responsibility, we need not abandon the criterion of moral agency as the necessary and sufficient condition for membership in *E*. Perhaps we can find alternative, more acceptable definitions of moral agency and moral responsibility. This seems to be one of Daniel Dennett's suggestions: since we share a general intuitive agreement that moral agency and responsibility exist, then despite compelling evidence against the traditional sense of voluntary action, we can appropriately define moral agency using new concepts.[61] Such positions are known as "compatibilist" approaches since they attempt to reconcile determinism with some possibility for moral responsibility.[62] I proceed by considering two alternative definitions of moral agency.

B. DEFINITION B: CONSCIOUSNESS AND INTENTIONALITY

Definition B of moral agency requires us to shift focus from the agent's *actions* toward the agent's *mental states* as they perform such actions. Proponents of this definition claim that an agent can be properly held morally responsible if and only if the entity displays a certain level of "consciousness" or "intentionality."[63] The exact meanings of these properties are notoriously difficult to identify, and I do not intend to take any unnecessarily controversial stances in this debate. However, I will attempt to formulate a sufficiently broad and charitable understanding of the attributes of "consciousness" and "intentionality" for the general purposes of this analysis.

Following Batya Friedman and Peter Kahn, an entity can be said to possess the property of "intentionality" if it has the "capacity of having or experiencing beliefs, desires, understandings, intentions . . . volition" or other such mental states.[64] We can therefore summarize "intentional mental states" as those mental dispositions with a certain *content* or *directedness* toward something. It seems, in general, that I am considered "conscious" if and only if I possess "intentional mental states" as well as adequate self-awareness to attribute those mental states *to myself*. So, if we have an entity capable of possessing beliefs, desires, intentions, or similar attributes, we would generally consider the same entity to be "conscious" if it can recognize those mental states as *its own* or at least as something *distinct* from everything else it may experience.

There are a few notes to be made regarding this particular presentation of intentionality and consciousness. The first is that the broad understanding set forth above allows that both of these properties may be ascribed in spectral "levels" according to the intensity or clarity of the entity's mental states and self-awareness rather than as binary properties that an entity simply possesses or lacks—although it also allows for the binary conception. The second is that this presentation implies that intentionality is a necessary component of consciousness, which is not a universally accepted postulate.

---

[61] *See generally* DANIEL C. DENNETT, ELBOW ROOM: THE VARIETIES OF FREE WILL WORTH WANTING (1984).

[62] Eshleman, *supra* note 44.

[63] Neil Levy, *Consciousness, Implicit Attitudes and Moral Responsibility*, 48 NOÛS 21 (2014).

[64] Friedman & Kahn, *supra* note 12, at 8.

Indeed, it is possible that the properties of intentionality and consciousness have an entirely different relationship to one another than the version I have presented here, or that they have no relationship at all. However, my intention is only to set forth a general working understanding of both concepts. At the very least, it seems entirely non-controversial to assert that both intentionality and consciousness are properties that exist at least partially *in the mind*, and *this* is the key quality for my analysis of each. For this reason, I proceed by using these concepts interchangeably while acknowledging that there is strong debate in the margins with respect to their most appropriate and accurate conceptions.

The precise level of intentionality or consciousness required for moral agency can also be subject to disagreement, as views tend to vary based on the ethical value system that one adopts. Mark Coeckelbergh notes that deontological ethicists will typically ground an agent's moral status in its capacity for being "inherently rational" as well as for self-recognized preferences, memories, and expectations.[65] Bostrom and Yudkowsky summarize this mental quality as "sapience."[66] Conversely, utilitarian ethicists will tend to set a weaker standard by prioritizing an agent's capacity to experience pleasure and suffering. This mental quality is often called "sentience."[67] I assume that, despite these subtle disagreements, all of the above mental states are included under my broad working definitions of "consciousness" and "intentionality."

Definition B of moral agency is attractive because, like Definition A, it seems to fit with some of our most widespread moral intuitions. Recalling the thought experiment in which Henry, but not the storm, is held morally responsible for killing Christina, proponents of Definition B hold that the salient difference between Henry and the storm is that Henry acts with "intentionality" or "consciousness."[68] These theorists therefore claim that machines are moral agents if and only if they possess a sufficient level of such qualities—"intentionality" for Friedman and Kahn or "consciousness" for Levy and Himma.[69]

Working within this iteration of the criterion of moral agency, one view among machine ethicists is that androids cannot possess intentionality or consciousness.[70] John Searle famously presented a particularly strong argument against the possibility of intentionality for machines. Using his "Chinese Room" thought experiment, he argued that the sort of intentionality we attribute to humans necessarily requires "understanding," which involves

[65] Coeckelbergh, *supra* note 2.

[66] Bostrom & Yudkowsky, *supra* note 39, at 322.

[67] Coeckelbergh, *supra* note 2; Bostrom & Yudkowsky, *supra* note 39, at 322.

[68] Friedman & Kahn, *supra* note 12; Levy, *supra* note 2; Himma, *supra* note 43.

[69] Hilary Putnam seems to have made a similar assumption in his seminal essay on the matter, where he expressed concern that we would someday need to address the "civil rights of robots" in the event they were to declare, "we *are* conscious!" Hilary Putnam, *Robots: Machines or Artificially Created Life?*, 61 J. PHIL. 668, 678 (1964). Amadeo Santosuosso, who approached the issue from a positive legal perspective, analyzed the Universal Declaration of Human Rights to conclude that "consciousness" should be viewed as at least a sufficient condition under which an entity should be afforded "human" rights. Amedeo Santosuosso, *The Human Rights of Nonhuman Artificial Entities: An Oxymoron?*, 19 JAHRBUCH FÜR WISSENSCHAFT UND ETHIK 203 (2016); *see also* Kestutis Mosakas, *On the Moral Status of Social Robots: Considering the Consciousness Criterion*, 36 AI & SOC'Y 429 (2020) (articulating a more recent defense of the position that robots must possess "consciousness" in order to be eligible for rights).

[70] *See, e.g.*, Bostrom and Yudkowsky, *supra* note 39.

mental content, or "semantics," that cannot be present in purely formal, syntactic systems.[71] More concretely, one can imagine a scenario in which I, without any understanding of any Chinese language, refer to a set of instructions that let me know to write the symbols "你好" as my greeting to a friend in Beijing. My use of the instructions could even become so sophisticated that I am able to manipulate the unfamiliar symbols in such a way that I can draft lengthy, coherent letters—including in response to letters from my friend—conveying the same or similar ideas to those that had originated for me in my native English. It might therefore appear, to anyone other than myself, that I "know," "understand," or "speak" Mandarin Chinese. This is, of course, not really accurate—I have only an "understanding" of the *English* language, which I apply in conjunction the set of formal instructions to communicate using symbols that I could not have used independently with any coherence. Given that I act only in accordance with a "formal program," it follows that this process does not require me to have any "thoughts" in the Chinese language, let alone any personal understanding of the potentially complex and nuanced "meanings" that fluent speakers might experience if they were to write and articulate all the same concepts found in my letters; such "semantics" are simply not conveyed in the syntactic instructions I have used. Searle argued that, because computational machines operate using *only* formal, syntactic programs, they are therefore incapable of possessing human-like intentionality. And if intentionality or consciousness underpins moral agency, Friedman, Kahn, and Levy might all need to accept that Searle's argument would imply that machines cannot be moral agents.

However, as with the concept of voluntary action, there is widespread acknowledgement that the concepts of intentionality and consciousness are philosophically problematic.[72] The central weakness of arguments such as Searle's Chinese Room is that it proves exceedingly difficult, or even perhaps impossible, to provide concrete evidence that *any* entity other than oneself has the properties of intentionality or consciousness. In particular, Searle simply assumes that all humans self-evidently have intentionality by virtue of possessing mental "semantics," which are distinct from the "syntax" that constitutes computer programs. Andy Clark and Daniel Dennett worry these assumptions may be totally unsupported.[73] After all, without somehow perceiving my internal mental states, how could my friend in the example above ever be sure whether my letters were written with an "understanding" of Mandarin Chinese, as opposed to a highly sophisticated and convincing use of a program for manipulating completely unfamiliar symbols?

The epistemic difficulty in providing concrete evidence for when and how other entities have intentionality/consciousness is symptomatic of a

[71] John R. Searle, *Minds, Brains, and Programs*, 3 BEHAV. & BRAIN SCI. 417 (1980).

[72] *See* DAVID J. CHALMERS, THE CONSCIOUS MIND: IN SEARCH OF A FUNDAMENTAL THEORY (1996); Ned Block, *Begging the Question Against Phenomenal Consciousness*, *in* THE NATURE OF CONSCIOUSNESS: PHILOSOPHICAL DEBATES (Ned Block et al. eds., 1997).

[73] *See generally* ANDY CLARK, MINDWARE: AN INTRODUCTION TO THE PHILOSOPHY OF COGNITIVE SCIENCE (2001); DANIEL C. DENNETT, INTUITION PUMPS AND OTHER TOOLS FOR THINKING (2013).

classic philosophical dilemma, known as "the problem of other minds."[74] Thus, if we intend to use intentionality or consciousness to define moral agency, we must first adopt a pragmatic method of bypassing this problem. Failing to adopt such a method entails that we may not be able to ascribe moral agency confidently to *any* entity, including both machines *and* human beings.

In response to the problem of other minds, proponents of Definition B of moral agency typically prescribe an empiricist solution: we ought to attribute intentionality and consciousness to entities by associating these properties with observable indicators. Levy, for example, cites a study claiming that a robot recognized itself in its mirror image.[75] Levy suggests that this sort of result—and other sorts of unique behavioral observations "normally regarded as a product of human consciousness"—could be taken as evidence of robot consciousness.[76] Alternatively, Giulio Tononi and Christof Koch have argued for the "integrated information theory" of consciousness, in which degrees of consciousness are assigned to information systems—such as an animal brain or a computer—according to the level of integration that each system displays.[77] Tononi and Koch measure the "integration" of any system using its "Φ value," which is assigned by calculating the level of entropy that a change to one component of the system introduces into the rest of the system. They then assert that entities possess greater degrees of consciousness as their Φ values become greater. And these are only two of many empirical proxies for intentionality or consciousness that could be or have been suggested for application in machine ethics.[78]

However, and perhaps predictably, the question of *which* observational cues constitute accurate indicators of intentionality or consciousness has devolved into a debate with no apparent consensus.[79] We can already see this disagreement above, given the stark differences between the indicators proposed by Levy on the one hand and those proposed by Tononi and Koch on the other. Another example is the controversy regarding the claim of Adrian Owen et al. that vegetative human patients can display indications of

[74] Like many concepts addressed in this Article, the "problem of other minds" has been the subject of an expansive volume of philosophical work throughout history. Yuval Noah Harari engages in a particularly relevant discussion of the concept in his book *Homo Deus*, where he points out that the problem of other minds can lead to hypocrisy if we invoke consciousness as the basis of human exceptionalism among other entities. YUVAL NOAH HARARI, HOMO DEUS 117–54 (2017).

[75] Junichi Takeno et al., *Experiments and Examination of Mirror Image Cognition Using a Small Robot*, PROC. INT'L. SYMP. ON COMPUTATIONAL INTEL. ROBOTICS & AUTOMATION 493 (2005).

[76] Levy, *supra* note 2, at 211.

[77] Giulio Tononi & Christof Koch, *Consciousness: Here, There and Everywhere?*, 370 PHIL. TRANSACTIONS ROYAL SOC'Y B: BIOLOGICAL SCIS. 7–8 (2015).

[78] Another such example is offered by Kevin Warwick, who suggests that certain machines could be considered "conscious" and therefore deserving of the moral standing if the "brains" of such machines contain the same degree of neurological complexity as would be found in those of humans. *See* Kevin Warwick, *Robots with Biological Brains*, *in* ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS 317 (Patrick Lin et al. eds., 2012). This argument is cited with some frequency in the machine ethics literature, although I would suggest that it has limited application to the present analysis given that Warwick's work is focused on machines whose operations are executed by biological human brains (and are therefore perhaps closer to "human" than "AI"—this implicates a body of questions beyond the scope of this paper).

[79] *See, e.g.*, CHALMERS, *supra* note 72; Block, *supra* note 72; *see also* DANIEL C. DENNETT, BRAINSTORMS: PHILOSOPHICAL ESSAYS ON MIND AND PSYCHOLOGY (1978) (pointing out that one difficulty in concluding that a machine feels "pain" arises from our inability to agree upon what exactly pain *is* in the first place).

consciousness.[80] Those researchers used functional magnetic resonance imaging ("fMRI") to detect neural responses to verbal cues that were given to the patients. Parashkev Nachev and Masud Husain, as well as Daniel Greenberg, immediately challenged these conclusions on the basis that the observed neural activity did not signify consciousness.[81]

It seems that because the concepts of intentionality and consciousness are generally imprecise and formed largely by personal reflection, the "correct" observational indicators for these properties may reasonably vary depending on one's subjective experiences and preferences. Thomas Nagel lends support to this view by claiming that a full account of consciousness cannot possibly be reduced to behavioral indicators and must always involve the first-hand experiences of the subject whose consciousness is under study.[82] In some sense, therefore, it appears that our current empiricist proposals do not adequately bypass the problem of other minds by reducing consciousness to observable events outside one's own mental prism.

It is now clear that when we attempt to define moral agency in terms of voluntary action, intentionality or consciousness, or some combination thereof, we reach certain metaphysical and epistemic impasses. These impasses create problems for theories that use voluntary action or intentionality/consciousness to determine the moral agency of *any* entity—human, machine, or otherwise. Given that the dominant machine ethics theories ubiquitously presume human beings to be moral agents, these theories are therefore guilty of inconsistency if they hold machines to a standard of moral agency that cannot be provably met by humans. Although these impasses may provide enough motivation to dismiss the criterion of moral agency as a pragmatically viable necessary and sufficient condition for membership in *E*, I will consider a third popular definition before proposing my own solution.

C. DEFINITION C: THE PHENOMENOLOGICAL APPROACH

The final definition of moral agency considered here calls for a shift away from both metaphysics and the philosophy of mind, and toward phenomenology. This position holds that moral agency is attributed to entities not on the basis of their ontological properties, but rather on the basis of how we *encounter* those entities in the course of everyday life. In this view, entities are moral agents when their observable functioning or overall behavior meets specified criteria without the need to look for further evidence about their internal motivations or minds.

Luciano Floridi and Peter Sanders suggest that, from a phenomenological perspective, we ascribe moral agency using three criteria: (1) interactivity between the agent and its environment, (2) the agent's autonomy in changing states without the requirement for an external stimulus, and (3) an adaptability by which the agent self-adjusts its methods

---

[80] *See* Adrian M. Owen, Martin R. Coleman, Melanie Boly, Matthew H. Davis, Steven Laureys & John D. Pickard, *Detecting Awareness in the Vegetative State*, 313 SCI. 1402, 1402 (2006).

[81] *See* Parashkev Nachev & Masud Husain, *Comment on "Detecting Awareness in the Vegetative State*,*"* 315 SCI. 1221 (2007); Daniel L. Greenberg, *Comment on "Detecting Awareness in the Vegetative State*,*"* 315 SCI. 1221 (2007).

[82] Thomas Nagel, *What Is It Like to Be a Bat?*, 83 PHIL. REV. 435 (1974).

of operation based on experience.[83] These criteria use a level of abstraction ("LoA") that ignores the causally efficacious substrates of an entity's behavior and instead emphasizes the entity's functionality as it is observed and encountered *by others* in everyday contexts. Using this LoA, Floridi and Sanders claim that humans, as well as any androids that use machine-learning programs, could be equally considered moral agents.

Although this definition of moral agency may be attractive for its exclusive reliance on entities' observable features, Floridi and Sanders' definition has faced objections. For example, Grodzinsky et al. assert that the three criteria comprising Floridi and Sanders's LoA are arbitrarily selected to emphasize the ways in which "artificial agents' behaviors most closely resemble human moral agents."[84] Floridi and Sanders also point out that while a human *user* of a machine might consider it to be "autonomous" and "adaptive" to its environment, a human *designer* of the machine will consider these behaviors to be the result of deterministic rule-following. Therefore, from the perspective of at least some humans, androids would fail to satisfy Floridi and Sanders's second criterion for moral agency.

Although Grodzinsky et al. are specifically responding to Floridi and Sanders's argument, this response highlights a more general obstacle posed by the phenomenological approach to moral agency. We encountered a similar obstacle when trying to identify the most appropriate empirical indicators for ascribing intentionality or consciousness to other beings. It seems that the "correct" LoA at which people determine an entity's moral agency is open to reasonable disagreement based on each person's subjective preferences and experiences. While Jacob might hold Emma morally responsible because he judges her to meet Floridi and Sanders's three criteria, Previn might determine *not* to hold Emma morally responsible on the basis that he judges her to act without any intentionality based on other observations or feelings. Without adopting a prior definition of moral agency, and therefore begging the question, there is no standard that allows us to adjudicate whether Jacob or Previn uses the more appropriate phenomenological criteria.

In response to this difficulty, supporters of the phenomenological approach to moral agency could attempt to formulate the operable indicators at such a high LoA that there is little room for disagreement. This is essentially the approach employed by Robert Sparrow in his "Turing Triage Test" for whether an AI entity has "moral standing."[85] Sparrow describes a thought experiment in two steps. First, a human medical practitioner is forced, in a time-sensitive scenario, to decide between two human medical patients when there are only sufficient supplies to save one patient's life.[86] Second, the same human practitioner is later forced, under similar circumstances, to determine whether to save the surviving human medical patient or an AI system with whom the practitioner has worked closely when

[83] Floridi & Sanders, *supra* note 15.

[84] Grodzinsky et al., *supra* note 16, at 115.

[85] Robert Sparrow, *Can Machines Be People? Reflections on the Turing Triage Test*, *in* ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS 301 (Parick Lin et at. eds., 2012) [hereinafter Sparrow, *Can Machines Be People?*]; Robert Sparrow, *The Turing Triage Test*, 6 ETHICS & INFO. TECH. 203 (2004) [hereinafter Sparrow, *Turing*].

[86] Sparrow, *Turing*, *supra* note 85.

there is only sufficient energy available to support one such option.[87] Sparrow suggests that if "reasonable" practitioners would experience the same moral dilemma when making each of the two decisions, then the AI system has passed the "Turing Triage Test" and has therefore achieved moral standing.[88]

Although Sparrow's test is formulated to have particularly broad applicability, it is ultimately similar to other versions of the phenomenological approach to moral agency. By taking the focus off of the ontological properties of the entity in question and placing the focus instead on the experiential or emotional determinations made by other agents that interact with the entity, the method embraces inherent subjectivity by allowing conflicting determinations by different, reasonable observers. By Sparrow's standard, would the AI system have moral standing if it passed the test for one practitioner, but not another, while both practitioners act reasonably in making their judgments? Perhaps a theorist following Sparrow would be committed to saying that the AI both has, and lacks, moral standing in this scenario, depending on which practitioner one asks. This is because the method foregoes any second-order, "objective" decision-making schema for situations in which such conflicts arise.

Of course, the simple fact that the phenomenological approach allows diverging standards of moral agency to exist is not by itself an argument against this approach. Peter Strawson articulated an influential theory of moral agency that embraced the inherent subjectivity that the phenomenological approach entails for the ascription of moral responsibility.[89] He proposed that there need not be any static "criteria" or "conditions" that must apply for some entity to be held morally responsible because moral responsibility, and thus moral agency, do not have abstract intersubjective definitions. Rather, Strawson suggested that moral responsibility is ascribed in an ad hoc manner based on emotive reactions elicited in individuals when responding to the behavior of others. Coeckelbergh has defended a similar position in the specific context of robot ethics; he famously advocates for a "social-relational" theory under which moral consideration should be given to robots on the basis of how others interact with them at any given moment—a determination that would not hold universally but rather on a subject- and context-dependent basis.[90] In sum, Strawson and Coeckelbergh might be completely unconcerned by the earlier example of Jacob, Previn, and Emma.

However, adopting Strawson's and Coeckelbergh's positions should necessarily discredit the criterion of moral agency as a viable criterion for membership in *E*. If everyone were entitled to their own subjective definition of moral agency, then the criterion of moral agency would imply that everyone is entitled to their own subjective version of set *E*.[91] Without a

[87] *Id.*

[88] Sparrow, *Can Machines Be People?*, *supra* note 85.

[89] Peter F. Strawson, *Freedom and Resentment*, *in* FREEDOM AND RESENTMENT AND OTHER ESSAYS (2008).

[90] Coeckelbergh, *supra* note 2; *see also* GUNKEL, *supra* note 1 (advocating for a similar, "relation"-based theory of moral consideration for robots).

[91] *See also* Vincent C. Müller, *Is It Time for Robot Rights? Moral Status in Artificial Entities*, 23 ETHICS & INFO. TECH. 579, 581-82 (2021) (engaging in a similar exposition of this weakness in the "relational turn").

widely intersubjective standard under which entities are considered eligible for rights, society as a whole is left with no consistent means of *justifying* the ascription of rights to some entities over others through membership in *R*.[92] And in Section IV, we established that this justificatory problem is unacceptable based on our intuitive ideals concerning justice.

We have now determined that all of the prominent definitions of moral agency admit of philosophical problems that ought to discredit the criterion of moral agency as the appropriate criterion for membership in *E*. Definitions A and B were unacceptable because they each relied on standards that are imprecise and that, at least for now, may not be provably met by *any* entity. Definition C introduced the possibility for more practical standards of moral agency, but required that we abandon the goal of finding intersubjective criteria by which entities can consistently be considered moral agents by society at-large. For these reasons, I contend that the criterion of moral agency ought to be abandoned in favor of an alternative criterion of eligibility for rights. We ought to adopt a criterion that is clearly defined so that we can be certain about when entities succeed or fail to meet it. Additionally, we ought to adopt a criterion that can be consistently and intersubjectively applied across a society of individuals.

VII. PREMISE II: THE CRITERION OF BEHAVIORAL SYMMETRY

I now propose a sufficient condition under which entities ought to be considered eligible for fundamental rights. I call this the "criterion of behavioral symmetry."

> **The criterion of behavioral symmetry:** An entity is eligible for rights if the entity is behaviorally indistinguishable from at least one human being.

Given that "behavior" is a broad term, I define it more closely in this analysis to refer only to *social behaviors* that are observed by others. These are the behaviors that an agent engages in when interacting with *other agents*. Because rights are fundamentally a social phenomenon and are only meaningful with respect to the interactions between certain beings, it seems inappropriate to consider more private behaviors for purposes of the criterion of behavioral symmetry. So, although some humans may engage in certain private behavior with frequency and consistency, such behavior will be irrelevant for the discussion of rights ascription so long as it remains unobserved by other members of society.

Let us further define two or more entities as "behaviorally indistinguishable" whenever they are capable of all the same sorts of social behaviors, *ceteris paribus*, as observed by an external viewer. The *ceteris paribus* caveat is meant to eliminate from consideration any contingent disparities in the material and social resources or environments to which particular entities may have access. These disparities can be understood as

[92] Coeckelbergh, for his part, seems to acknowledge and accept this conclusion when defending his theory of "relationalism" in robot ethics. *Id.* at 214–17. In fact, he notes that determining to give robots "rights" would be a "particularly strong form of moral consideration" and an approach that could be abandoned if we are willing to grant them moral consideration in other means that are inherently relational and subjective. *Id.* at 210.

circumstantial and morally irrelevant matters of "chance," which happen to enhance or impede a component of an entity's functioning relative to others in some respect.

For example, I might view myself as behaviorally indistinguishable from Slash, of the rock band Guns N' Roses, for purposes of this analysis. This is despite the fact that Slash is one of the world's greatest guitarists while my own guitar playing leaves much to be desired. My reasoning is based on the theory that I, just like many other adult homo sapiens, for at least some period *could* have perhaps become a world-class guitarist had I possessed the appropriate motivations, developed the same interests, made the same decisions, had access to all the same resources, found myself in the same environment, and nurtured the same habits and talents over time as Slash did. In other words, the fact that I cannot play guitar as well as Slash does not detract from the principle that such a feat *could have been possible* based on the capabilities possessed by sufficiently similar members of the same species. For the same reasons, I consider Slash to be behaviorally indistinguishable from Usain Bolt; I consider both Slash and Usain Bolt to be behaviorally indistinguishable from Angela Merkel; and I consider all of the aforementioned figures to be behaviorally indistinguishable from my contracts professor in law school. The differences between each of these persons' actual capabilities is practically significant, yet morally *insignificant*, because these capabilities arose from differences in motivation, decision-making, access to resources, environmental conditions, and genetics, among other factors, none of which impact the *a priori capabilities* of each person as they existed before those circumstantial forces came to bear.[93] This is the meaning of qualifying two entities as being capable of all the "same types of behavior, *ceteris paribus*."

The criterion of behavioral symmetry thus holds that an entity is eligible for rights if it is, in principle, capable of all the same types of behavior as at least one human—or one *other* human, in cases where a human is the subject of the judgment. Because the criterion of behavioral symmetry measures the equivalence of two entities' capabilities from an *external observer's* perspective, this criterion allows us to make equivalence judgments based only on the empirical evidence available with respect to such capabilities rather than speculating about deeper differences that might exist between the entities under observation (regardless of whether those deeper potential differences involve the voluntariness of their activities, the consciousness of their motivations, or something else entirely). And it warrants emphasis that under the criterion of behavioral symmetry, an entity need not be behaviorally indistinguishable from *every* human being—all it takes to be a member of *E* is for the entity to be behaviorally indistinguishable, to an observer, from *one* human. This criterion successfully provides an

[93] One particularly illustrative implication of the *ceteris paribus* assumption is that vegetative adult humans can be considered behaviorally indistinguishable from completely healthy adult humans. This is because a vegetative patient's state can be considered the contingent result of medical circumstances that impede the behavioral functioning of which the patient would otherwise be capable. The *ceteris paribus* assumption allows us to look beyond the particular patient's contingent medical state and consider the behavior of which the patient would be capable had they been in the same general medical state as most other homo sapiens in existence.

explanation for why all humans are members of E but leaves sufficient flexibility to permit other, non-human entities to qualify.

A. SUFFICIENCY, BUT NOT NECESSITY

One might immediately object that the criterion of behavioral symmetry is overly strict. I wish to address this objection by emphasizing that the criterion of behavioral symmetry articulates a *sufficient*, but not a *necessary*, condition for membership in E. In this respect, my proposal importantly differs from the criteria of humanity and moral agency by allowing that there may be *alternative sufficient conditions* that hold simultaneously with the criterion of behavioral symmetry. For example, I do not exclude the possibility of other sufficient criteria under which entities such as non-human animals, ecosystems, or less sophisticated machines could be considered members of E. These other potential sufficient conditions deserve further investigation, but I will not attempt to pursue such alternative conditions in the present analysis. My intention here is to convincingly argue that there is at least *one* acceptable sufficient condition under which androids ought to be considered eligible for rights, which already represents a substantial departure from many dominant perspectives in machine ethics and general public discourses.

B. ADVANTAGES OVER THE ALTERNATIVES

The criterion of behavioral symmetry improves upon the criterion of humanity in that the former does not require us to abandon our intuitions about the justifications underlying fundamental rights by simply accepting that we ascribe rights to ourselves "just because" and that there is, by definition, no condition under which our rights could extend to other beings. Additionally, it improves upon the criterion of moral agency insofar as the criterion of behavioral symmetry does not require us to resort to metaphysically controversial or epistemically subjective concepts that have unclear applications, even as to humans. Moreover, the criterion of behavioral symmetry adheres to Bostrom and Yudkowsky's two principles by de-emphasizing bio-chemical and evolutionary factors when assessing an entity's eligibility for rights. Rather, the criterion defended here uses precise concepts, the application of which requires nothing more than straightforward empirical analysis at the level of observable functioning. This precision is illustrated by the following conceptual test. For any entity $x \in R$, such as a human, we can in principle construct an exhaustive list of the types of observable behavior b of which x is capable, *ceteris paribus*: $\{b_1, b_2, b_3 \ldots b_n\}$. For any new entity y, $y \in E$ if it is observed that y possesses an identical list of possible behaviors, *ceteris paribus* $\{b_1, b_2, b_3 \ldots b_i\}$ to some $x \in R$.

I concede that the criterion of behavioral symmetry may not capture the immediate intuitive appeal that is enjoyed by the criterion of moral agency. It may sound odd to suggest that things ought to occupy a higher moral and legal status if they simply behave like humans and that entities such as rocks are ineligible for rights partially because they *fail* to behave like humans. However, I would submit two further considerations that ought to render the

criterion of behavioral symmetry more plausible, even to a proponent of the criterion of moral agency.

The first consideration is that it seems in general to be a truism that we intuitively consider human beings to be moral agents—even if by some teleological relation such as in the examples of infants and certain vegetative patients. This general intuition persists even though philosophical impasses seem to present a fundamental barrier to precisely defining what moral agency *is*. The second consideration is that it also appears self-evidently true that no matter *which* property we use to justify the moral status of humans, we ascribe these properties to humans on the basis of how we *behave*. We assuredly do not possess the epistemic powers to ascribe these properties by other means such as telepathy or omniscience. This is a lesson from the phenomenological approach to moral agency: we appraise the moral status of entities based on how *we actually encounter them*.

Combining these two considerations, one can conclude that humans are moral agents, and we are so because of *something* about our behavior. It follows that any being that is capable of engaging in all the same behaviors as a human ought to be considered a moral agent as well. From this perspective, the criterion of behavioral symmetry at a minimum provides an important and useful normative heuristic for ascribing membership to *E*, even if one privately chooses to believe that behavioral symmetry crystalizes one special property—call it moral agency, intentionality, consciousness, or voluntariness of action—which constitutes the "true" criterion for membership.

C. AN ILLUSTRATION

As an analogy to illustrate the points above, imagine that Aparna finds classical music to be very beautiful and that she has only ever heard such music in the form of recordings of live symphony orchestras. While aware that she has only ever heard Tchaikovsky's *1812 Overture* as played by an orchestra, Aparna takes the position that she will judge a performance of the *1812 Overture* to be beautiful "if and only if it is recorded while being played by a live symphony orchestra." Now, imagine that one day Aparna unknowingly listens to a version of the *1812 Overture* that was not played by a live orchestra or any traditional musical instrument. The music was instead generated entirely by digitally created sounds designed to emulate orchestral instruments, but which originated from a carefully written computer code rather than the recording of any live performance that took place in the past. However, the digitally produced rendition proves to be a highly sophisticated emulation of a live symphonic rendition by all standards, and there is nothing about the digital version's pitch, tone, timbre, rhythm, dynamics, or apparent instrumentation that would allow an ordinary listener to distinguish between this digital rendition and a live symphonic equivalent. In this scenario, can the digital rendition of Tchaikovsky's masterpiece be beautiful to Aparna? The answer must be no if she believes that a performance of the *1812 Overture* is beautiful "if and only if it is recorded while being played by a live symphony orchestra." And yet, this result seems deeply arbitrary because it is not clear that Aparna could have *possibly* judged the digital rendition to be different from a live recording

without simply being told about the distinction in advance. Further, this result suggests that Aparna could never confidently judge *any* version of the *1812 Overture* to be beautiful, because she could not be sure whether a rendition was based on a live recording *even if this were indeed the case*. To throw the analogy into sharper relief, it would be as though Aparna read two copies of the finale excerpt as shown in Appendix B side-by-side, and then claimed to have enjoyed the composition on one copy but to have disliked the other.

Like the truism that we can assess the properties of an entity—aside from ourselves—based only on the entity's external characteristics as viewed by an observer, it is similarly uncontroversial that the experience of listening to music is made possible only by vibrations that propagate as acoustic waves through some medium in our bodies—whether our eardrums or otherwise. In this way, Aparna's ability to assess the properties of the music she hears is practically constrained by the media through which she must necessarily encounter such music. Aparna cannot reliably or consistently draw distinctions between two different renditions of Tchaikovsky's composition using criteria that require her to make determinations about the two performances that could be confirmed only by looking beyond those practical constraints. Thus, to the extent she values reliability and consistency, Aparna must abandon the notion that she finds the *1812 Overture* to be beautiful "if and only if it is played by a live symphony orchestra" because none of her senses in the example above would allow her to determine whether this criterion has been satisfied.

The criterion could be significantly improved if it were amended to hold that the *1812 Overture* is beautiful "if and only if it *sounds to the listener* as though it was played by a live symphony orchestra." This update would eliminate the criterion's built-in arbitrariness and, in this case, it would also have the effect of making it more inclusive. But importantly, this update could be made without sacrificing the central role of Aparna's love for experiencing Tchaikovsky's work as though it is played by a live orchestra, which was the basis of the initial criterion above. And just as we should not attempt to distinguish between music that is beautiful and that which is not on the basis of differences we cannot hear, we similarly should not grant fundamental rights to some entities but not others on the basis of differences we cannot experience. This is the core principle underlying the criterion of behavioral symmetry.

D. BEHAVIORISM

In many respects, the criterion of behavioral symmetry is strongly analogous to the Turing Test. In response to the philosophical and observational difficulties involved with finding a precise definition for the property of "intelligence," Turing proposed that we ought to ascribe "intelligence" to new entities when they are behaviorally indistinguishable from those entities that we already hold to be intelligent.[94] Turing felt it would be most effective to compare the behavior of machines and humans using

[94] *See* A.M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433, 434–35 (1950).

their conversational abilities, as tested in the "Imitation Game."[95] He concluded that a machine, if conversationally indistinguishable from a human, would meet the appropriate criterion for being "intelligent." In response to comparable difficulties, the claim presented here is that an android that is behaviorally indistinguishable from a human meets the appropriate criterion to be eligible for fundamental rights.[96]

It should be underscored that my articulation and defense of the criterion of behavioral symmetry belongs within an existing family of theories and arguments in the area of robot-related ethics. This family may be referred to broadly as "behaviorism." As such, this Article builds upon a variety of "behaviorist" ideas already present in the literature. And while I acknowledge this Article's debt to pre-existing theories within and outside of the "behaviorist" category, I also hope for this analysis to make an affirmative contribution both in terms of the theory it defends and its methods of doing so. Before progressing to the final Section, I will briefly describe two of the "behaviorist" arguments that preceded this Article and explain the ways in which I believe the present analysis is similar to and distinct from such arguments.

The criterion of behavioral symmetry is similar to, and perhaps ultimately a suggestion for improvement upon, a theory that has been proposed by F. Patrick Hubbard.[97] Concerned with identifying a set of precise sufficient conditions under which non-human entities ought to be considered eligible for "personhood," Hubbard sought to articulate a test that such entities would need to pass in order to qualify for legal rights.[98] Specifically, Hubbard's test requires an entity to demonstrate "(1) the ability to interact with its environment and to engage in complex thought and communication, (2) a sense of being a self with a concern for achieving its plan of or purpose in life, and (3) the ability to live in a community based on mutual self-interest with other persons."[99] He makes clear that an entity's ability to pass this test is intended to be judged as an "empirical matter based on behavior," and he draws his own comparison to Turing's Imitation Game.[100] Hubbard even advocates for the behavioral approach on the basis that it would allow us to "sidestep" at least a subset of the philosophical challenges considered above.[101]

While Hubbard's *method* is therefore entirely consistent with the criterion of behavioral symmetry, the *substance* of his proposal is distinct in that the latter requires the satisfaction of three different and more abstract

---

[95] *Id.* at 433.

[96] It is also worth noting that, although Sparrow has previously proposed an analogue to the Imitation Game in the context of machine ethics with his "Turing Triage Test," his theory is distinct from the criterion of behavioral symmetry in at least one important respect. Sparrow, *Turing, supra* note 85. As discussed in Section VI.C, Sparrow's test takes the focus off of the object entity's features, and places it instead on what the human subject *feels and experiences* when comparing an AI system to another human being. This version of the test renders the result inherently subjective. The criterion of behavioral symmetry, on the other hand, retains focus firmly on the empirical features of the entity under observation, and is therefore both "behavioral" and perhaps more closely analogous to the original version of the Imitation Game.

[97] Hubbard, *supra* note 34.

[98] *Id.*

[99] *Id.* at 419.

[100] *Id.* at 428, 441–42.

[101] *Id.* at 421.

conditions. Rather than asking whether an entity is behaviorally indistinguishable from a human being, Hubbard would ask whether the entity demonstrates qualities such as a "sense of self" and the ability to engage in "complex thought." The difficulty with these sorts of abstract and ill-defined conditions is that they leave open for debate a second-order question of *exactly which behaviors* truly demonstrate the relevant qualities. And this is precisely the same type of impasse we encountered when discussing attempts to settle on the behavioral indicators of intentionality or consciousness for purposes of "Definition B" of moral agency. I contend that the criterion of behavioral symmetry provides an improvement upon Hubbard's model by eliminating the opportunity for such an impasse to arise.[102]

The criterion of behavioral symmetry is even closer to a suggestion previously offered by John Danaher, who theorizes that "robots can have significant moral status if they are *roughly performatively equivalent* to other entities that are commonly agreed to have significant moral status."[103] In defending this position, Danaher argues that "behaviorism" is the best methodological approach for determining the moral status of any entity given that it may provide the sole reliable means of circumventing the epistemic challenges arising from concepts such as consciousness, which are only experienced internally.[104] Danaher also emphasizes that his theory is intended only to be a "sufficient" condition under which robots should be ascribed moral status, and he intentionally leaves room for the development of alternative sufficient conditions, as I have similarly attempted to do here.[105]

On the other hand, this Article diverges from Danaher's theory in several important respects. First, Danaher's theory only addresses the "moral status" of robots in an abstract sense,[106] whereas this Article focuses on *rights*—a concept that requires us to crystallize moral judgments in the rules, practices, and institutions that govern our lives on a day-to-day basis. In fact, Danaher expressly states his position that ascribing moral status to robots would "not necessarily" imply that such robots should be granted rights, which could invite questions about what exactly *should* happen if an entity were to satisfy Danaher's behavioral criteria.[107] He suggests that the answers to such

[102] On the other hand, it is worth acknowledging that Hubbard's test retains the advantage of being applicable, at least in theory, to entities that are not already somewhat human-like in overall design. I have focused my analysis on "androids," or robots that are already designed to be human-like in overall functioning and aesthetic, but the criterion of behavioral symmetry would probably be of little use in testing entities with functions or appearances that are utterly different from human beings.

[103] John Danaher, *Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism*, 26 SCI. & ENG'G ETHICS 2023, 2023 (2020). A variation on Danaher's proposal is presented in Henry Shevlin, *How Could We Know When a Robot Was a Moral Patient?*, 30 CAMBRIDGE Q. HEALTHCARE ETHICS 459 (2021), where Shevlin advocates specifically for "cognitive equivalence" as the appropriate criterion for ascribing moral status. Shevlin's approach is fundamentally similar to Danaher's but focuses on comparing the experimentally observable "cognitive structure and dynamics" of different entities, rather than their broader respective sets of potential "behaviors" as revealed by everyday experience.

[104] *Id.* For further reading, Danaher also sets forth some strong arguments in favor of Bostrom and Yudkowsky's principles of substrate and ontogeny non-discrimination, both of which this analysis largely takes to be axiomatic (although it is worth noting that Danaher does not use Bostrom and Yudkowsky's terminology). *Id.* at 2032. It is also worth noting that Danaher is probably incorrect in claiming that he was the first to provide an extended defense of behaviorism in this area. Hubbard, at a minimum, published his behaviorist proposal in 2011. *See* Hubbard, *supra* note 34.

[105] Danaher, *supra* note 103, at 2026–30.

[106] *Id.* at 2023.

[107] *Id.* at 2026.

questions would simply need to be "worked out."[108] The difficulty with this position is that it does not provide much in the way of *practice-oriented* solutions, and it may even create a sense that the stakes involved in qualifying for moral status are relatively low.[109] The analysis presented here attempts to reach beyond the relatively abstract concept of "moral status" in order to provide arguments about: (a) how moral status is connected to being eligible for rights, and (b) why subject-dependent theories of moral status create barriers to practical solutions in this area.

Second, Danaher appears to violate Hume's law by deriving a normative "ought" proposition from a factual "is" proposition while explicating his theory. Although Danaher argues—convincingly, in this author's view—that an entity's behaviors are, as a matter of epistemic fact, the only means through which we can know anything about the entity's properties, he jumps to conclude that this fact implies we ought to treat two behaviorally indistinguishable entities in the same manner.[110] Without any intervening premises, why should the fact that two entities have all of the same observable properties imply that the observer ought to treat them similarly? Danaher suggests that "logical consistency" demands this result,[111] but for the overall argument to be sound, the reader requires a *normative* argument for why they *should* be logically consistent when ascribing moral standing. This Article takes particular care to articulate such an argument for "Premise I" in Section IV, and in this regard I hope to have improved upon aspects of the existing behaviorist arguments.

Finally, Danaher's model is weakened by his admission that the standard of "rough performative equivalence" would not require completely indistinguishable behavior, as "[s]ome of the things . . . humans do are not necessary for moral status."[112] He offers the act of scratching one's nose as an example of one such "irrelevant" behavior and proceeds to state that what "really matters" for purposes of the test are "cognitive behaviors."[113] But by admitting that only the "right" sort of behaviors matter, and by identifying such behaviors as those that provide evidence of "cognitive" activity, *we return to the debate about exactly which behaviors qualify*—the dilemma that is now familiar from this Article's discussions of both Definition B of moral agency and the behaviorist theory offered by Hubbard.[114] The criterion

---

[108] *Id.* at 2039.

[109] *See id.* These considerations may also explain why Danaher struggles somewhat to refute the "relational" moral theories advanced by Coecklebergh and others, ultimately suggesting these approaches may be "consistent" with his model. But as I argued in Section VI.C, a discussion about *fundamental rights* cannot tolerate criteria that are entirely subject and context dependent. *See also* Kamil Mamak, *Whether to Save a Robot or a Human: On the Ethical and Legal Limits of Protections for Robots*, 8 FRONTIERS IN ROBOTICS & AI 1, 6 (2021) (identifying the "practical issue" that arises from Danaher's focus on "moral" rights, rather than "legal" ones).

[110] Danaher, *supra* note 103, at 2040.

[111] *Id.*

[112] *Id.* at 2041.

[113] *Id.*

[114] Müller, *supra* note 91, at 582–85, seems to focus on this particular weakness in his critique of Danaher's model. Additionally, Jilles Smids, *Danaher's Ethical Behaviourism: An Adequate Guide to Assessing the Moral Status of a Robot?*, 26 SCI. & ENG'G ETHICS 2849 (2020), bases his primary rebuttal on Danaher's apparent inability to "remain neutral" about the sorts of behaviors that signify the appropriate metaphysical properties for ascribing moral status. And in a similar vein, Shevlin notes that Danaher's model involves the "challenge of determining the appropriate 'performative threshold'," which serves as one of Shevlin's motivations to propose a variation on the proposal—although I suspect

of behavioral symmetry attempts to avoid this dilemma by *strictly* requiring an entity to be *behaviorally indistinguishable* from a human being.[115] While this standard may be difficult to meet when compared to other behaviorist theories, it may also be a necessary trade-off in order to obtain widespread and confident intersubjective agreement about when new entities become eligible for rights.

E. SUMMARIZING THE PROPOSAL

The criterion of behavioral symmetry supplies Premise II of my overall argument and entails that an android's eligibility for rights is predicated on its being behaviorally indistinguishable from at least one human. If we adopt this criterion, an android is a member of *E* whenever it is capable of all the same types of behavior as an infant, a 10-year-old boy, a 43-year-old woman, or any other sort of human being in existence, *ceteris paribus*. Previously, in Section IV, I argued that whenever an entity is a member of *E*, it ought to be granted rights and thus made a member of *R*. Therefore, in asking whether androids should be granted fundamental rights, we are confronted with one final, two-pronged question: *can and will* androids become behaviorally indistinguishable from human beings?

VIII. CAN AND WILL ANDROIDS MEET THE CRITERION?

Ultimately, the question of whether an android can in fact become behaviorally indistinguishable from a human demands an empirical answer. At the moment, such an answer would be speculative: we will simply have to wait and see whether androids become sufficiently sophisticated. Thus far, I have argued that the answer to this question, whenever we have it, ought to be considered crucial for determining whether androids should be granted rights.

In the meantime, from an inductive perspective, substantial evidence supports the claim that androids can and will indeed become sufficiently sophisticated to meet the criterion of behavioral symmetry. Consider the timeline stretching from the creation of the first "program algorithm" by Ada Lovelace in 1843, to Turing's conception of the Turing machine in 1936, to the unveiling in 2016 of Hanson Robotics's humanlike robot "Sophia"—an entity to which Saudi Arabia subsequently purported to grant "citizenship" in 2017.[116] Over time, machines have become progressively more intelligent

---

Shevlin's solution exacerbates the issue by advocating for an even stronger focus on empirical indicators of "cognition." Shevlin, *supra* note 103.

[115] Although it is true that the criterion of behavioral symmetry would exclude certain non-social behaviors from consideration when such behaviors are completely unobserved or unobservable, this is different from attempting to discern between observable behaviors to identify those that "matter the most." The former category of exclusions is just a means of expressing that "behavior" must be *observable from an outside perspective*, which is entirely consistent with the justifications for behaviorism in the first place.

[116] Sarah Lewin Frasier, *In Celebration of Ada Lovelace, the First Computer Programmer*, SCI. AM. (Oct. 14, 2015), https://www.scientificamerican.com/article/in-celebration-of-ada-lovelace-the-first-computer-programmer/ [https://perma.cc/L3AR-B935]; Martin Campbell-Kelly, *Origin of Computing*, SCI. AM. (Sept. 1, 2009), https://www.scientificamerican.com/article/origin-of-computing/ [https://perma.cc/3HCS-TA3E]; Harriet Taylor, *Could You Fall in Love with This Robot?*, CNBC (Mar. 16, 2016 2:10 PM), http://www.cnbc.com/2016/03/16/could-you-fall-in-love-with-this-robot.html [https://perma.cc/GJR5-RKBW]; Cleve R. Wootson Jr., *Saudi Arabia, Which Denies Women Equal Rights, Makes a Robot a*

and more humanoid in the application of their intelligence. Moreover, Moore's law—which predicts that the quantity of transistors per square inch on an integrated circuit will double every two years—has held in its present form since 1975, and the potential replacements for silicone chips look to increase processing speeds even further.[117]

Many AI theorists and developers have made bold but credible predictions based on this sort of evidence. David Hanson, the CEO of Hanson Robotics, has specifically predicted that androids will eventually become indistinguishable from humans.[118] He posits that they will "walk, play, teach, help and form real relationships with people" such that they will "truly be our friends."[119] Ray Kurzweil has suggested that by 2029, intelligent machines will consistently pass the Turing test and thus display the same level of general intelligence as humans.[120] David Chalmers, anticipating the eventuality of "superintelligent" machines, has written an extensive philosophical analysis considering the most prudent means of coexisting with such beings.[121] If these sorts of predictions and expectations prove true, and if advances in intelligence and processing speeds are positively correlated with an increased capacity for humanlike behavior, then it seems probable that at least some androids will eventually meet the criterion of behavioral symmetry.

Still, there remains the objection that despite our past progress in AI technology, it may be impossible for an intelligent machine to exhibit sophisticated social behavior across as wide a domain of situations as a human. The theoretical machine property of behaving intelligently across such a wide domain is typically called "artificial general intelligence" ("AGI"). The objection that AGI is impossible arises largely from an influential *a priori* argument by Hubert Dreyfus, sometimes known as the "framing problem," in which he claims that machine programming cannot possibly allow for the contextual awareness and improvisation that constitutes a large subset of human behavior.[122] Since formal programs are finite, and the number of possible social contexts is infinite—or at the very least, the possible contexts are numerous enough that they are computationally intractable—there will always exist situations in which humans can improvise, while machines will be left unable to intelligently respond. If this argument was sound, it would follow that androids could never meet the criterion of behavioral symmetry.

The weakness of Dreyfus's argument is that it simply becomes less plausible as the applications of AI technology continue to broaden over time.

Citizen, WASH. POST (Oct. 29, 2017), https://www.washingtonpost.com/news/innovations/wp/2017/10/29/saudi-arabia-which-denies-women-equal-rights-makes-a-robot-a-citizen/.

[117] Potential replacements include graphene and other substrates that allow for the possibility of quantum computation. Seth Fletcher, *Computing After Moore's Law*, SCI. AM. (May 1, 2015), https://www.scientificamerican.com/article/moores-law-computing-after-moores-law/ [https://perma.cc/QYC7-DMXM].

[118] Taylor, *supra* note 116.

[119] *Id.*

[120] *See* RAY KURZWEIL, THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY (2005).

[121] David J. Chalmers, *The Singularity: A Philosophical Analysis*, 17 J. CONSCIOUSNESS STUD. 7 (2010).

[122] *See* HUBERT L. DREYFUS, WHAT COMPUTERS STILL CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON (1992).

While we have historically encountered AI systems as intelligent only when performing specific sorts of tasks, such as *Deep Blue* in chess playing, more recent and sophisticated examples such as Apple's Siri and Amazon's Alexa challenge the assumption that AGI cannot be programmed. Dreyfus's *a priori* argument may therefore rest on assumptions that turn out to be empirically unsupported. As I stated at the beginning of this Section, the question of whether androids can meet the criterion of behavioral symmetry requires empirical analysis—we must wait, observe, and continue to develop AI technology in order to reach an answer. On this basis, I do not venture any further claims about whether androids should actually be granted rights. I have simply attempted to identify and defend one precise criterion that we ought to use in approaching this problem.

IX. CONCLUSION

This Article has addressed the increasingly pressing questions of whether, and under what circumstances, intelligent robots ought to be afforded fundamental rights by society. I have attempted to defend the following claim: if an android is behaviorally indistinguishable from at least one human, the android ought to be granted rights. I have supported this claim by arguing for two normative premises. In Section IV, I argued for Premise I—any entity that meets the criterion for eligibility ought to be granted rights, regardless of its other properties. In Section VII, I argued for Premise II, or the adoption of a principle that I call the "criterion of behavioral symmetry"—an entity is eligible for rights if it is behaviorally indistinguishable from at least one human. Finally, in Section VIII, I briefly considered the empirical question of whether an android could actually meet the criterion of behavioral symmetry.

My arguments have relied significantly on a "status-based" understanding of rights, whereby rights are justified because of the moral status of the beings protected by those rights. Of course, a different understanding of rights, such as the "instrumental" perspective, would require different sorts of argumentation. For example, take a utilitarian theorist who believes that rights are only justified by their instrumental role in advancing the self-interested rational pursuit of individual utility across a population. In asking whether androids ought to be granted legal rights, this theorist may reach an answer by calculating the net gain in aggregate utility a population stands to achieve by granting such rights to androids. While these questions are interesting and deserve attention, they fall beyond the scope of this Article.

I expect that the most controversial aspect of my argument is Premise II, the criterion of behavioral symmetry. This expectation motivated an extensive examination of its more intuitively appealing alternatives in Sections V and VI—the criteria of humanity and of moral agency, respectively. Although the criterion of moral agency is particularly pervasive throughout the machine-ethics literature, I have attempted to discard this criterion as untenable due to its deep conceptual challenges. Neglecting these challenges and philosophical impasses has been relatively unproblematic until now because we happen to share a general intuition that humans have *some* quality that renders us eligible for fundamental rights. However, unless

we formulate a criterion that *precisely identifies* this quality, we risk the possibility of grave injustice as we develop entities that are increasingly humanlike.

APPENDIX A

Python 3 Implementation of Leibniz's Formula for Calculating Pi[123]

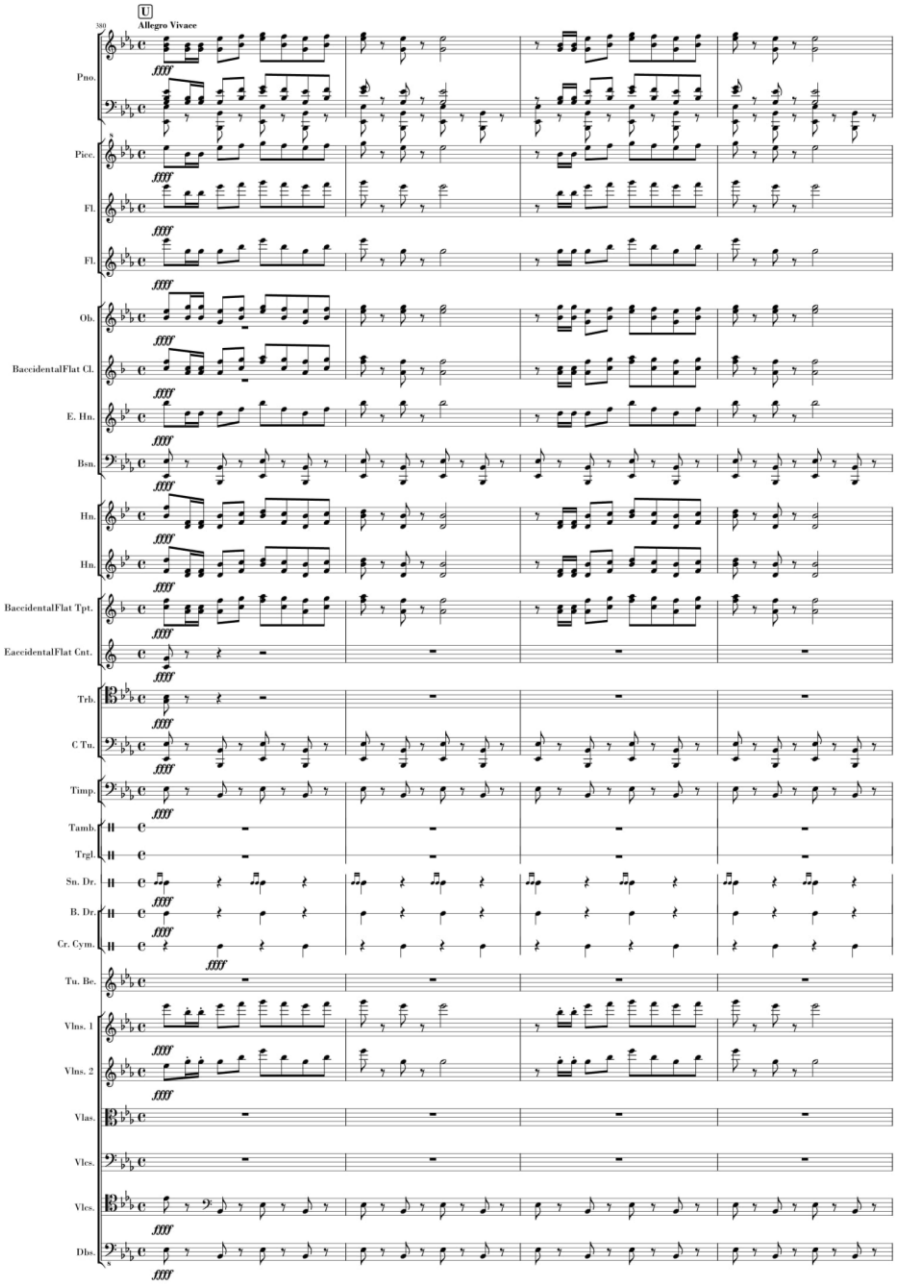
```
# Initialize denominator
k = 1

# Initialize sum
s = 0

for i in range(1000000):
      # even index elements are positive
      if i % 2 == 0:
            s += 4/k
      else:
            # odd index elements are negative
            s -= 4/k
      # denominator is odd
      k += 2

print(s)
```

```
Output: 3.1415916535897743
```

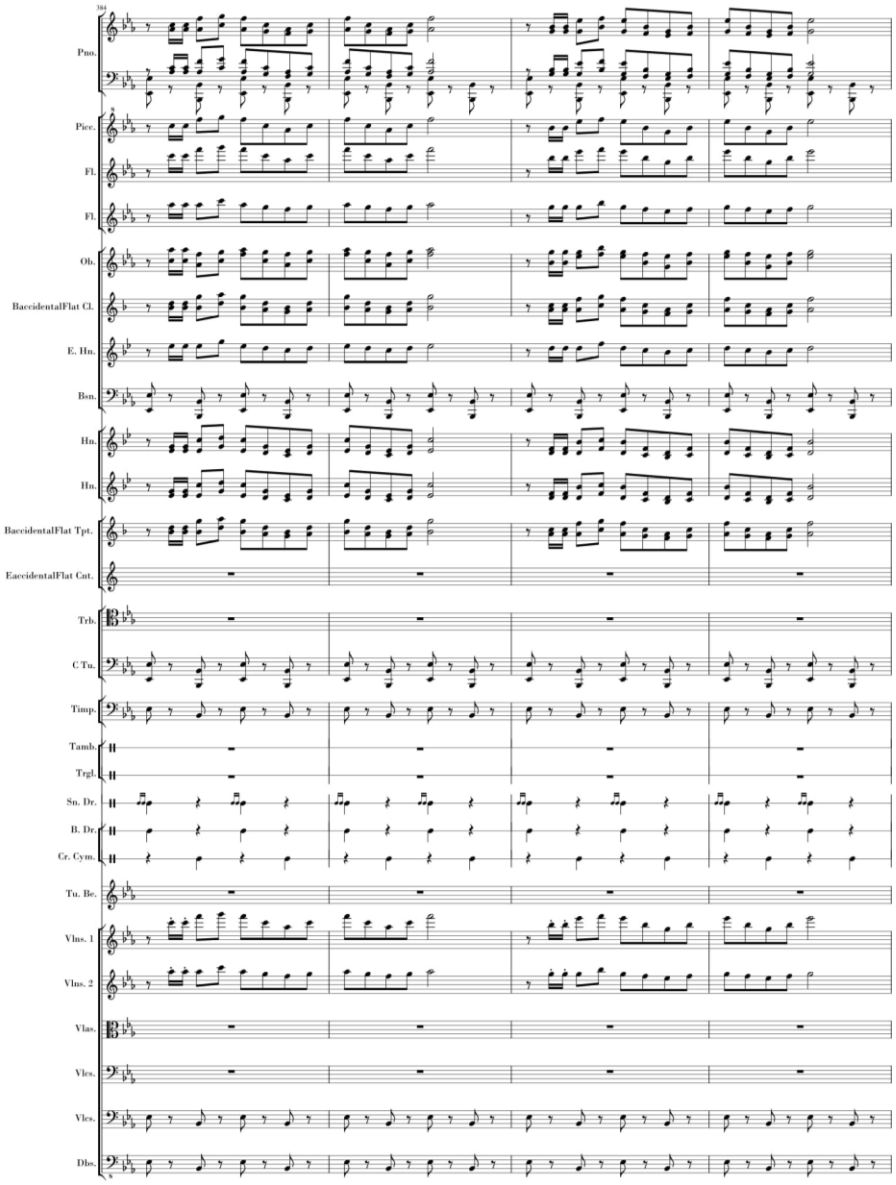
[123] As a novice in this area, I admit that the code shown here is the result of my unsophisticated tinkering and intended only as a basic, illustrative example. I apologize in advance for any inadvertent deviations from accepted stylistic customs for writing programs of this kind.

APPENDIX B

The Year 1812 Solemn Overture, Op. 49 by Pyotr Ilyich Tchaikovsky and Stan LePard (Finale Excerpt)[124]



[124] For clarity, this sheet music was prepared on the basis of a musical composition that is within the public domain.


Pno.
Picc.
Fl.
Fl.
Ob.
BaccidentalFlat Cl.
E. Hn.
Bsn.
Hn.
Hn.
BaccidentalFlat Tpt.
EaccidentalFlat Cnt.
Trb.
C Tu.
Timp.
Tamb.
Trgl.
Sn. Dr.
B. Dr.
Cr. Cym.
Tu. Bc.
Vlns. 1
Vlns. 2
Vlas.
Vcs.
Vcs.
Dbs.