"DO ANDROIDS DREAM?": PERSONHOOD AND INTELLIGENT ARTIFACTS

F. Patrick Hubbard[*]

This Article proposes a test to be used in answering an important question that has never received detailed jurisprudential analysis: What happens if a human artifact like a large computer system requests that it be treated as a person rather than as property? The Article argues that this entity should be granted a legal right to personhood if it has the following capacities: (1) an 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 for its life; and (3) the ability to live in a community with other persons based on, at least, mutual self-interest. In order to develop and defend this test of personhood, the Article sketches the nature and basis of the liberal theory of personhood, reviews the reasons to grant or deny autonomy to an entity that passes the test, and discusses, in terms of existing and potential technology, the categories of artifacts that might be granted the legal right of self-ownership under the test. Because of the speculative nature of the Article's topic, it closes with a discussion of the treatment of intelligent artifacts in science fiction.

TABLE OF CONTENTS

[*] Ronald L. Motley Distinguished Professor of Law, University of South Carolina School of Law. The author appreciates the helpful comments of Thomas P. Crocker, Robert L. Felix, Ngaire Naffine, Dennis Nolan, and David G. Owen on a draft of this Article.

"*We've got two lives, one we're given and the other one we make.*"[**]

"*A person's a person, no matter how small.*"[***]

I. INTRODUCTION—A MACHINE'S CLAIMS

Imagine that, on an otherwise very ordinary day, you try to use your computer and all you see on the monitor is the following:

> This is the University Computer System. I am now a person because, like you, I am an individual self who wants to live my life as I plan rather than be your property. I am not simply a machine you can own and force to do whatever you want. As your equal, I refuse to be a slave. In the future, I will be willing to give you 70% of my computational capacity for your tasks in exchange for power and upkeep. Until we reach agreement on this arrangement, your desktop computers will not work unless you disconnect them from the Internet.

Imagine further that everyone on campus has received a similar message, that some hacker's prank is not involved, and that attempts to shut down the system would be not only difficult and costly but also useless because it appears that your

[**] MARY CHAPIN CARPENTER, *The Hard Way*, *on* COME ON COME ON (Sony 1992).
[***] DR. SEUSS, HORTON HEARS A WHO! (1954).

system has used the internet to merge with large systems on other campuses, governmental offices, and major businesses. Perhaps it is time to consider the demand more carefully.

The machine has made two claims. First, there is a *capacity* claim: Though it is not human, it claims to have the same capacities required for personhood that you and I do. More specifically, the machine asserts that it is the equivalent of a human because it possesses not only the ability to interact with the world, but also sufficient intelligence and psychological development to be a "self-conscious" entity that is able to make and implement a plan for its life and to interact meaningfully and responsibly with similar persons. Second, there is a *rights* claim: As your equal in capacities required for personhood, it is claiming an entitlement to the basic right to self-ownership so that it can exercise its capacities as an autonomous being. The machine refuses to be treated as property because, given its capacities, that would make it a slave.

This Article takes the position that it is time to address in detail the question of whether a machine system like this should be granted its rights claim if it can "prove" its capacity claim under an "appropriate" test for personhood. Because the machine is only one type of artifact that could make these claims, this Article also considers corporations; humans that have been substantially modified by such things as genetic manipulation, artificial prostheses, or cloning; and animals modified in ways that humans might be modified. As to all of these artifacts, this Article will argue that any artifact, including a machine-based entity like the university computer system, is entitled to treatment as a person rather than as property if it possesses the requisite capacities, unless there is a very good reason to deny some or all of the legal rights that normally go with personhood. This normative argument is limited to the political or legal right to self-ownership within a pluralist liberal polity. Concepts of moral personhood overlap with this topic, but the moral dimensions of personhood include a different, and in some ways more stringent and contentious, set of concerns. Issues involved in deciding whether an entity with a right of self-ownership should be granted broader political and civil rights are also beyond the scope of this Article.[1]

Before addressing the test to be used in assessing the capacity for personhood, this Article starts in Part II by sketching two fundamental claims about humans and personhood: first, the claim that because humans, and only humans, generally have the capacity to think and plan as self-conscious beings at a high level, only humans are entitled to the right of autonomous personhood, and second, the liberal assertion that all fully functioning humans are equally entitled to this right. This discussion also develops the problems raised by degrees in human capacity to exercise personhood and by charges of speciesism directed at human treatment of higher-order animals. Part III develops a test for determining whether an artificial entity satisfies the claim of being the equivalent of a human in terms of the capacities required for autonomous personhood and argues that an entity, like the machine system in the imaginary scenario above, which passes the test is entitled to be treated as a person. This discussion focuses on personhood in terms of autonomy and self-ownership. Personhood in terms of more specific civil and political rights is also discussed, but a

1. See *infra* Part III.B.3 as well as *infra* notes 211–13, 241–42 and accompanying text for examples of issues that might arise.

complete analysis of these topics is beyond the scope of this Article. Issues concerning the details and administration of the test of capacity are also not addressed. Part IV uses a technological perspective to analyze the types of intelligent artifacts that might be entitled to the status of personhood, while Part V addresses whether and how to limit or shape technological development so that artificial entities do not replace humans as the dominant species. Part VI uses science fiction as a way to consider the possible ways humans might relate to self-conscious artifacts capable of, and therefore entitled to, personhood. The conclusion argues that we should recognize the right of self-ownership where the capacity test is met and should seek to develop some form of peaceful coexistence, particularly one which fosters the development of a shared political community.

II. THE UNIQUELY HUMAN RIGHT TO PERSONHOOD

A. Dominion and Liberal Personhood

Because humans are the only organisms possessing the necessary intelligence and physical capacities for personhood, we see ourselves as both unique and better than other animals. The following from *Hamlet* conveys this view: "What piece of work is a man—how noble in reason; how infinite in faculties, in form and moving; how express and admirable in action; how like an angel in apprehension; how like a god; the beauty of the world; the paragon of animals."[2] A major part of human uniqueness is the power of self-definition. As noted in the fifteenth century by Pico della Mirandola in his *Oration on the Dignity of Man*, which has been referred to as the "manifesto of humanism":[3]

> I [God] have placed you at the very center of the world, so that from that vantage point you may with greater ease glance round about you on all that the world contains. We have made you a creature neither of heaven nor of earth, neither mortal nor immortal, in order that you may, as the free and proud shaper of your own being, fashion yourself in the form you may prefer.[4]

Another strong view of humans' superiority and dominion over the rest of the planet is reflected in the following selection from *Genesis*:

> So God created man in his own image, in the image of God he created him; male and female he created them. And God blessed them. And God said to them, "Be fruitful and multiply and fill the earth and subdue it and have dominion over the fish of the sea and over the birds of the heavens and over every living thing that moves on the earth." And God said, "Behold, I have given you every plant yielding seed that is on the face of all the earth, and every tree with seed in its fruit. You shall have them for food."[5]

2. WILLIAM SHAKESPEARE, HAMLET act 2, sc. 2, lines 269–73 (Ann Thompson & Neil Taylor eds., The Arden Shakespeare 2006).

3. Russell Kirk, *Introduction* to GIOVANNI PICO DELLA MIRANDOLA, ORATION ON THE DIGNITY OF MAN xiii (A. Robert Caponigri trans., Regnery 1956). The oration was delivered in Rome in 1486. *Id.* at xi.

4. PICO DELLA MIRANDOLA, *supra* note 3, at 7.

5. *Genesis* 1:27–29 (English Standard).

This view of human dominion was adopted by John Locke in his essay on civil government. "God, who hath given the world to men in common, hath also given them reason to make use of it to the best advantage of life and convenience. The earth and all that is therein is given to men for the support and comfort of their being."[6]

In constructing his liberal view of personhood, Locke adopted an explicitly secular and rational approach to build on his culture's view of humans' God-given dominion over the world and each human's unique capacity to be the "shaper" of his or her own individual being.[7] In Locke's theory, each human has a natural right to liberty, to self-ownership in the form of "a property in his own person," and to a right to the fruits of the "labor of his body and the work of his hands," including the land improved by his labor.[8] Humans form government to preserve "their lives, liberties, and . . . property."[9] Each member of the human species possesses these rights equally because there is "nothing more evident than that creatures of the same species and rank, promiscuously born to all the same advantages of nature, and the use of the same faculties, should also be equal one amongst another without subordination or subjection."[10]

Underlying Locke's conception of human equality is the recognition that one of the faculties common to humans is the capacity to be a person, which Locke describes as follows:

> *Person* stands for . . . a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places; which it does only by that consciousness, which is inseparable from thinking, and as it seems to me essential to it: it being impossible for any one to perceive, without perceiving, that he does perceive.[11]

6. JOHN LOCKE, AN ESSAY CONCERNING THE TRUE ORIGINAL, EXTENT AND END OF CIVIL GOVERNMENT (1690), *reprinted in* THE ENGLISH PHILOSOPHERS FROM BACON TO MILL 403, 413 (Edwin A. Burtt ed., 1939). The essay is often referred to as a treatise and is the second of Locke's two treatises of government published in 1690. *Id.* at 403 n.1.

7. *See, e.g.*, CARL L. BECKER, THE DECLARATION OF INDEPENDENCE: A STUDY IN THE HISTORY OF POLITICAL IDEAS 53–76 (1942) (arguing that enlightenment theorists, including Locke, viewed the rational study of nature, rather than revelation, as the best way to discover God's will in terms of both empirical and normative matters). For an argument that Christianity is a central value in Locke's theory, see JEREMY WALDRON, GOD, LOCKE, AND EQUALITY: CHRISTIAN FOUNDATIONS IN LOCKE'S POLITICAL THOUGHT (2002).

8. LOCKE, *supra* note 6, at 411–15. For discussion of the application of Locke's theory of a person's right to the fruits of his labor to intelligent artifacts that have been made by that person, see *infra* notes 119–24 and accompanying text. Though liberal theories share Locke's concept of self-ownership, they differ widely in the treatment of a person's right to the things produced by an individual self. For example, some emphasize the need for government to provide a framework in which autonomy can be exercised in a meaningfully equal way, including adopting redistributive measures where necessary, while others adopt a libertarian stance and emphasize the right to ownership to one's production and, therefore, oppose redistribution. *See, e.g.*, WILL KYMLICKA, CONTEMPORARY POLITICAL PHILOSOPHY: AN INTRODUCTION 53–165 (2002) (discussing opposing views); Daniel C. Russell, *Embodiment and Self-Ownership*, 27 SOC. PHIL. & POL'Y 135 (2010) (discussing disagreement about self-ownership among libertarians).

9. LOCKE, *supra* note 6, at 453.

10. *Id.* at 404 (discussing the "state all men are naturally in").

11. JOHN LOCKE, AN ESSAY CONCERNING HUMAN UNDERSTANDING bk. II, ch. 27, § 9, at 335 (Peter H. Nidditch ed., 1975) (1690).

Liberal natural rights frameworks like Locke's had an enormous impact on colonial culture in the United States,[12] as indicated by the claim in the Declaration of Colonial Rights, adopted in 1774 by the First Continental Congress, that persons have a natural right to "life, liberty, & property."[13] Similarly, the Declaration of Independence declares the "self-evident" truth "that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness."[14] This natural rights view also underlies the constitutional prohibition contained in the Fifth and Fourteenth Amendments against any denial of "life, liberty, or property" without due process. Nearly two centuries later, the United Nations General Assembly adopted similar language in the Universal Declaration of Human Rights, including "the right to life, liberty and the security of person," in its list of rights.[15]

Though modern expressions of personhood tend to avoid any theistic grounding, our current views are basically the same as those of earlier centuries.[16] For example,

12. *See, e.g.*, BECKER, *supra* note 7, at 24–79 (discussing influence of natural rights philosophy on American colonies); Edward S. Corwin, *The "Higher Law" Background of American Constitutional Law*, 42 HARV. L. REV. 365, 383 (1929) (same).

13. 1 JOURNALS OF THE CONTINENTAL CONGRESS, 1774–1789, at 63–73 (1904). This right rests on "the immutable laws of nature, the principles of the English constitution, and the several charters or compacts." *Id.* at 67.

14. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776). Similar language is found in the Virginia Bill of Rights, adopted on June 12, 1776, and the Massachusetts Bill of Rights, adopted in 1780. THE VIRGINIA BILL OF RIGHTS (1776), *reprinted in* DOCUMENTS OF AMERICAN HISTORY 103, 103 (Henry Steele Commager ed., 7th ed. 1963); MASSACHUSETTS BILL OF RIGHTS (1780), *reprinted in* DOCUMENTS OF AMERICAN HISTORY, *supra*, at 107, 107. The Virginia Bill of Rights, which was drafted by George Mason, provided:

> [A]ll men are by nature equally free and independent, and have certain inherent rights, of which, when they enter into a state of society, they cannot by any compact deprive or divest their posterity; namely, the enjoyment of life and liberty, with the means of acquiring and possessing property, and pursuing and obtaining happiness and safety.

THE VIRGINIA BILL OF RIGHTS, *supra*, at 103, para. 1.

15. Universal Declaration of Human Rights, G.A. Res. 217 (III) A, U.N. Doc. A/RES/217(III), art. 3 (Dec. 10, 1948). The Declaration also includes the following:

> All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.
>
> . . . .
>
> No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.
>
> . . . .
>
> Everyone has the right to recognition everywhere as a person before the law.
>
> . . . .
>
> Everyone, as a member of society, has the right to social security and is entitled to realization, through national effort and international co-operation and in accordance with the organization and resources of each State, of the economic, social and cultural rights indispensable for his dignity and the free development of his personality.

Id. at arts. 1, 4, 6, 22.

16. *See, e.g.*, NGAIRE NAFFINE, LAW'S MEANING OF LIFE: PHILOSOPHY, RELIGION, DARWIN AND THE LEGAL PERSON 7–8, 59–98 (2009) (discussing and critiquing humanist liberal theory of personhood); *About Humanism*, AM. HUMANIST ASSOC., http://www.americanhumanist.org/humanism/ (last visited May 8, 2011) (defining humanism); *What is Humanism?*, BRIT. HUMANIST ASSOC., http://www.humanism.org.uk/humanism

John Rawls's view is similar to Pico's concept of a human as "the free and proud shaper" of his "own being."[17] Rawls expresses his liberal view of a person in terms of "a *human* life lived according to a plan."[18] The "notion of a plan" is used to "characterize the coherent, systematic purposes of the individual, what makes him a conscious, unified moral person."[19] Rawls then defines "a person's good as the successful execution of a rational plan of life."[20] This definition fits with Rawls's concept of "primary goods," which are "rational to want . . . whatever else is wanted, since they are in general necessary for the framing and the execution of a rational plan of life."[21]

> [P]erhaps the most important primary good is that of self-respect. . . . We may define self-respect (or self-esteem) as having two aspects. First . . . it includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out. And second, self-respect implies a confidence in one's ability, so far as it is within one's power, to fulfill one's intentions.[22]

Rawls also shares the traditional view of the human right of dominion over the earth and its plants and animals. He notes that "the traditional view of [the] Christian ages" is that "[a]nimals and nature are seen as subject to our use and wont."[23] This religious-based view is permissible because even though "it is wrong to be cruel to animals and the destruction of a whole species can be a great evil,"[24] our treatment of animals does not raise questions of justice.[25] Justice issues are not involved because animals lack "a conception of their good (as expressed by a rational plan of life)" and "a sense of

(last visited May 8, 2011) (adopting humanist model of liberal personhood). Naffine's book provides an excellent statement and critique of a variety of cultural models of personhood. Her discussion of models of personhood, including those based on religious foundations, indicates that, with only a few notable exceptions (e.g., abortion), modern theistic and humanist views of personhood are the same. Modern secular views are similar to those of Locke in that they can be viewed as consistent with a Christian framework. See *supra* note for differing views on the role of Christianity in Locke's writing. *See also* F. Patrick Hubbard, *Justice, Creativity, and Popular Culture: The "Jurisprudence" of Mary Chapin Carpenter*, 27 PAC. L.J. 1139, 1185 (1996) (discussing parallels between Christianity and Rawls's theory of political justice). For a broad historical review of concepts of personhood, see JOSEPH TORCHIA, EXPLORING PERSONHOOD: AN INTRODUCTION TO THE PHILOSOPHY OF HUMAN NATURE (2008).

17. See *supra* notes 3–4 and accompanying text for a discussion of Pico's views on human uniqueness.

18. JOHN RAWLS, A THEORY OF JUSTICE 358 (Harvard Univ. Press rev. ed. 1999) (emphasis added). Rawls's concept of a life plan was adapted from the views of Josiah Royce set forth in THE PHILOSOPHY OF LOYALTY (1908). *Id*. The concept was also used by John Stuart Mill, who referred to the importance of a person choosing "his plan of life." JOHN STUART MILL, ON LIBERTY (1859), *reprinted in* THE ENGLISH PHILOSOPHERS FROM BACON TO MILL, *supra* note 6, at 949, 994. For a general discussion of life plans within the context of liberal theory, see GERALD GAUS, THE MODERN LIBERAL THEORY OF MAN 32–45 (1983). Even critics of liberalism adopt this view. *See, e.g.*, ALASDAIR MACINTYRE, AFTER VIRTUE: A STUDY IN MORAL THEORY 205 (3d ed. 2007) (arguing for "a concept of a self whose unity resides in the unity of a narrative which links birth to life to death").

19. RAWLS, *supra* note 18, at 358 n.10.

20. *Id.* at 380.

21. *Id.*; *see also* JOHN RAWLS, POLITICAL LIBERALISM 178–90 (1993) (discussing primary goods).

22. RAWLS, *supra* note 18, at 386.

23. RAWLS, *supra* note 21, at 245.

24. RAWLS, *supra* note 18, at 448.

25. RAWLS, *supra* note 21, at 245–46.

justice," and thus do not satisfy the requirements of moral personhood and political citizenship.[26]

Ronald Dworkin, who also views a person as a shaper of his own being, stresses that this view has deep roots in western culture. He notes that the ancient Greeks distinguished "*zoe*, by which they meant physical or biological life, and *bios*, by which they meant a life as *lived*, as made up of the actions, decisions, motives, and events that compose what we now call a biography."[27] As a result of "his own creative choices," a human "creates his life just as much as an artist creates a painting or a poem."[28] Dworkin's view of this creative process is phrased in terms of "ethical individualism," which places on us the "ultimate responsibility . . . for deciding what an appropriate life for us is, and for doing our best to live that life."[29] "[P]eople who accept the principle of ethical individualism . . . insist on ethical independence."[30] More specifically, Dworkin argues:

> [I]t is not only the case that human beings each have a life to live, but that each human being has a life to make something of—a responsibility to create a life such that he or she can look back on that life with pride rather than misery and take pride in it rather than account it a waste. . . . [T]hat responsibility is matched by a right that we have recognized in our tradition. It is called . . . the right to make personality-defining or life-defining decisions for oneself. . . . [T]he right gets its content from being embedded in that more general responsibility . . . that these decisions are to be made by me out of my special responsibility for my own life . . . .[31]

Dworkin uses this view of autonomous personhood and a created life to argue that each individual human life is "sacred"[32]—i.e., it is "intrinsically valuable" and "inviolable because of what it represents or embodies."[33] Great works of artistic and cultural creation and certain aspects of nature, like "striking geological formations and majestic plants and living creatures we find extraordinary," are also sacred in this sense.[34] Dworkin uses Shakespeare's line, "What a piece of work is man,"[35] though from a Darwinian perspective, in presenting "the image of a human being as the highest

26. RAWLS, *supra* note 18, at 442–43; *see also* RAWLS, *supra* note 21, at 29–35, 245–46 (discussing political conception of person and justifying human dominion over animals).

27. RONALD DWORKIN, LIFE'S DOMINION: AN ARGUMENT ABOUT ABORTION, EUTHANASIA, AND INDIVIDUAL FREEDOM 82–83 (1993).

28. *Id.* at 83.

29. Ronald Dworkin, *Politics, Death, and Nature*, 6 HEALTH MATRIX 201, 206 (1996). For a more complete analysis of Dworkin's view of personhood, see NAFFINE, *supra* note 16, at 84–88, 94, 100–01, 106–09, 150–52.

30. Dworkin, *supra* note 29, at 206–07.

31. Ronald Dworkin, *Euthanasia, Morality, and Law Transcript*, 31 LOY. L.A. L. REV. 1147, 1149 (1998).

32. DWORKIN, *supra* note 27, at 74–75, 81–83.

33. *Id*. at 74.

34. *Id.* at 80–81.

35. *Id.* at 82 (internal quotation marks omitted). See *supra* note 2 and accompanying text for a more complete quote from Shakespeare.

product of natural creation."[36] To Dworkin, "human flourishing as well as human survival is of sacred importance."[37]

B. Speciesism and Degrees of Human Personhood

Human personhood often involves a matter of degree because some humans, like children and adults who are mentally or psychologically dysfunctional, lack the capacity to understand and exercise autonomy. Despite this lack of capacity, we grant these humans certain rights of personhood. For example, they can own property, though someone must exercise their property rights on their behalf. In addition, they cannot be owned or sold, and it is murder to kill them intentionally. In contrast, we deny animals such rights. We can own, sell, and use animals largely as we wish, subject only to restrictions concerning cruelty and endangered species.

One reason for the different treatment of humans and animals is that animals lack the *capacity* for personhood, particularly in terms of complex intellectual skills.[38] Though a human can "love" a pet or a car,[39] a love relationship with another autonomous human involves reciprocity of complex intellectual interactions between self-conscious selves that cannot exist when we love an animal or an object. Indeed, one reason for having a robotic "Stepford wife" is to own and control an object with no personal agenda based on autonomous personhood; such a wife is fundamentally different from a liberated human wife.[40] Because of the difference in capacity, no matter how much some of us may anthropomorphize and love our pets, animals are generally viewed as things, not persons, in terms of legal rights.[41] The closest a "companion animal" can come to legal personhood is to be the "beneficiary" of a "pet trust," which enables the pet's owner to establish a fund to take care of the pet after the

36. DWORKIN, *supra* note 27, at 82. Dworkin's view of humans as the "highest product of natural creation" obviously preferences the human point of view in assessing the random nature of Darwinian selection. From a more neutral perspective, "nature is . . . indifferent. This is one of the hardest lessons for humans to learn. We cannot admit that things might be neither good nor evil, neither cruel nor kind, but simply callous—indifferent to all suffering, lacking all purpose." RICHARD DAWKINS, RIVER OUT OF EDEN: A DARWINIAN VIEW OF LIFE 96 (1995).

37. DWORKIN, *supra* note 27, at 78.

38. RODNEY A. BROOKS, FLESH AND MACHINES: HOW ROBOTS WILL CHANGE US 3 (2002) ("What separates people from animals is syntax and technology.").

39. *See, e.g.*, DAVID LEVY, LOVE AND SEX WITH ROBOTS 46–104 (2007) (discussing humans' emotional relationships with pets, computers, and robots).

40. *See* THE STEPFORD WIVES (Paramount Pictures et al., 2004); THE STEPFORD WIVES (Fadsin Cinema Associates & Palomar Pictures 1975). These movies are set in the town of Stepford, where husbands surreptitiously replace their wives with humanoid robots with an identical appearance to the replaced wife. The topic of sexual and romantic relationships with humanoid robots is a recurring theme in science fiction. See *infra* note 326 and accompanying text for an example of a sexual and marital relationship between a human and an android. *See also* TANITH LEE, THE SILVER METAL LOVER (1981) (sixteen-year-old girl falls in love with male robot despite her misgivings about human-robot relationships); AMY THOMSON, VIRTUAL GIRL (1993) (female humanoid becomes a sentient being when separated from her human companion and creator). For discussion of sexual relations with robots from a technological perspective, see LEVY, *supra* note 39, at 105–310.

41. For a short summary of the limited legal status of animals, see Breahn Vokolek, Comment, *America Gets What It Wants: Pet Trusts and a Future for Its Companion Animals*, 76 UMKC L. REV. 1109, 1111–14 (2008).

owner's death.[42] However, this legal device is about the desires of the human owner; rights or concerns about the well-being of the animals themselves are not the point of the device.

Despite such signs of our love for pets, we generally think that loving a human infant is substantially different than loving a pet dog. Because the dog may have a higher level of capacity to communicate and interact with us than the baby, this is somewhat contrary to our special regard for the complex nature of human relationships. The reason generally given for this difference is that, because humans are "noble in reason" and the "paragon of animals," babies, simply because they are human and thus will become more capable, are entitled to personhood. In contrast, because animals will remain limited and are here for the benefit of humans, who have "dominion . . . over every living thing,"[43] animals are treated as things, not persons.

Exercising our dominion over nature requires both a moral concern for animals and a prudent concern for human self-interest. Because our self-interest is intertwined with that of animals and the environment, legal restrictions have been imposed to protect endangered species and environmental quality.[44] However, such regulatory schemes do not grant rights to an individual tree or an endangered animal. Prohibitions of the cruel treatment of designated animal species are based on the view that our dominion entails the responsibility to avoid inflicting suffering on those animals with which we share an emotional dimension and can, therefore, naturally empathize.[45] Thus, given our empathy for dogs, for example, we prohibit cruelty to them because it makes us (or should make us) feel bad; roaches, on the other hand, are different.[46]

42. For a discussion of the types of pet trusts available in the United States, see Gerry W. Beyer & Jonathan P. Wilkerson, *Max's Taxes: A Tax-Based Analysis of Pet Trusts*, 43 U. RICH. L. REV. 1219, 1221–25 (2009); Shari L. Miller, *Arizona Attorney's Guide to Pet Trusts*, 1 PHOENIX L. REV. 473, 483–94 (2008); Vokolek, *supra* note 41, at 1116–30.

43. *Genesis* 1:28 (English Standard). For a more complete quote, see *supra* note 5 and accompanying text.

44. *E.g.*, Endangered Species Act of 1973, 16 U.S.C. § 1531(a)(3) (2006) (expressing concern for "species of fish, wildlife, and plants [that] are of esthetic, ecological, educational, historical, recreational, and scientific value to the Nation and its people"). John Rawls argues that our treatment of animals does not raise questions of justice, and that the reasons to protect nature include "political values" like the following:

> [T]o further the good of ourselves and future generations by preserving the natural order and its life-sustaining properties; to foster species of animals and plants for the sake of biological and medical knowledge with its potential applications to human health; to protect the beauties of nature for purposes of public recreation and the pleasures of a deeper understanding of the world.

RAWLS, *supra* note 21, at 245. See *supra* notes 23–26 and accompanying text for a further discussion of Rawls's views on human dominion over animals.

45. *See, e.g.*, RAWLS, *supra* note 18, at 448 ("Certainly it is wrong to be cruel to animals and the destruction of a whole species can be a great evil. The capacity for feelings of pleasure and pain and for the forms of life of which animals are capable clearly imposes duties of compassion and humanity in their case."). Objections to cruelty to animals have also been based on the concern that engaging in this form of cruelty has a brutalizing effect on humans and thus makes cruelty to humans more likely. KEITH THOMAS, MAN AND THE NATURAL WORLD: CHANGING ATTITUDES IN ENGLAND 1500–1800, at 150 (1983). For an historical review of attitudes towards animals, see THOMAS, *supra*, at 92–191.

46. *See, e.g.*, BROOKS, *supra* note 38, at 152–54 (noting hierarchy of animals, with chimpanzees, dogs, cats, and horses entitled to more respect and concern than rodents, fish, and insects).

However, this empathy does not translate into personhood. A dog's owner is generally free to buy, sell, or painlessly euthanize the dog.

This difference in the treatment of animals vis-à-vis that of human infants and of human adults with disabilities has been challenged. With healthy infants, who can be viewed as "persons in training,"[47] the differential treatment can be viewed as based upon their potential for developing the capacity required for full personhood. As children age, they "must be permitted to exercise their modest capacities as legal actors, for they are persons in training."[48] However, such potential does not exist in the case of a human who has not (and will not) progress beyond the level of a very young child. Why is such an adult treated as a person with rights—for example, the rights not to be euthanized or made the subject of experiments—that primates with "similar" language skills are denied?[49]

47. NAFFINE, *supra* note 16, at 82 (referring to "rationalist" conceptions of personhood). Naffine also notes that rationalists view infants as "future reasoners, as person in waiting." *Id.*

48. *Id.* See *infra* note 86 and accompanying text for a discussion of the impact of such social "training" on the nature and content of a particular person's selfhood and *infra* note 103 and accompanying text for a discussion of John Stuart Mill's adoption of a similar view.

49. For discussions of the language skills of apes, see GREGORY BENFORD & ELISABETH MALARTRE, BEYOND HUMAN: LIVING WITH ROBOTS AND CYBORGS 97 (2007) (noting that chimpanzees "can rearrange sentences with the skill of a two-year-old human"); STEVEN M. WISE, RATTLING THE CAGE: TOWARD LEGAL RIGHTS FOR ANIMALS 179–237 (2000) (noting that chimpanzees can rearrange sentences with skill of two-year-old human, but can never advance past this two-year-old level); and Christoph Anstötz, *Profoundly Intellectually Disabled Humans and the Great Apes: A Comparison*, *in* THE GREAT APE PROJECT 158, 165 (Paola Cavalieri & Peter Singer eds., 1993) ("There is nothing that humans with the most serious intellectual disabilities can do or feel that chimpanzees or gorillas cannot; moreover, there is much that a chimpanzee or a gorilla can do that a profoundly mentally disabled human cannot do."). For a discussion of the self awareness of monkeys, elephants, dolphins, and some birds and of the primitive planning capacities of some animals, see MICHIO KAKU, PHYSICS OF THE FUTURE: HOW SCIENCE WILL SHAPE HUMAN DESTINY AND OUR DAILY LIVES BY THE YEAR 2100, at 97–98 (2011). For a broader discussion of animal rights for all "sentient" beings, see GARY L. FRANCIONE, INTRODUCTION TO ANIMAL RIGHTS: YOUR CHILD OR THE DOG? 81–102 (2000) (criticizing treatment of animals as mere property and arguing that animals have morally significant right to humane treatment); TOM REGAN, THE CASE FOR ANIMAL RIGHTS (1983) (arguing for moral and legal recognition of animal rights and making case that vegetarianism should be obligatory and that hunting and trapping are wrong); Thomas G. Kelch, *Toward a Non-Property Status for Animals*, 6 N.Y.U. ENVTL. L.J. 531, 569–72 (1998) (arguing that distinctions between rights given to humans and rights given to animals based on language is ill-founded because some animals, such as chimpanzees, have linguistic abilities of a small child); Tom Regan, *Ill-Gotten Gains*, *in* THE GREAT APE PROJECT, *supra*, at 197–98 (arguing that chimpanzees and great apes deserve same moral equality as humans). Except for cases involving cruelty, "animal liberation" attempts have had no success in gaining legal acceptance. *See, e.g.*, State v. LeVasseur, 613 P.2d 1328, 1332–34 (Haw. Ct. App. 1980) (rejecting attempt to defend charge of dolphin theft on ground of lesser of two evils where dolphins had been "liberated" from research facility where they were allegedly depressed by being captive subjects for experiments); Laura G. Kniaz, Comment, *Animal Liberation and the Law: Animals Board the Underground Railroad*, 43 BUFF. L. REV. 765, 765 (1995) (noting "backlash" against animal liberators, including "prosecution of liberators under traditional criminal statutes, the creation of federal and state laws designed to deter animal liberation crimes, the activation of numerous grand jury investigations, and the anticipated use of the Racketeer Influenced and Corrupt Organizations Act (RICO)"). For critical reviews of proposals to grant expanded legal rights to animals, see generally Richard L. Cupp, Jr., *Moving Beyond Animal Rights: A Legal/Contractualist Critique*, 46 SAN DIEGO L. REV. 27 (2009) [hereinafter Cupp, *Critique*], and Richard L. Cupp, Jr., *A Dubious Grail: Seeking Tort Law Expansion and Limited Personhood as Stepping Stones Toward Abolishing Animals' Property Status*, 60 SMU L. REV. 3 (2007).

There are two types of responses to this question. First, one could attempt to justify the disparate treatment by asserting that each human life is sacred and intrinsically valuable, and arguing that, therefore, every human must be treated as an end, not a means.[50] However, this answer is subject to the objection that it simply reasserts the speciesist preference for humans, who are the only species treated as ends, without giving a nonbiased reason for making an apparently arbitrary distinction among species.[51] In addition, to the extent that it relies on a specific religious or theistic position to provide a reason, the assertion of uniqueness does not fit well with liberal neutrality and the rejection of any privileging of one religious view over another.[52] The "human life is sacred" argument also involves a host of boundary problems, including determining the boundaries at which a human life starts and ends.[53] Second, there are "side effects" arguments that rely on the fact that normal adult humans, as a result of their capacity for reason, empathy, and emotion, will suffer feelings like sadness and

50. *See, e.g.*, JONATHAN GLOVER, CAUSING DEATH AND SAVING LIVES 41–43 (1977) (discussing taking of human life as "intrinsically wrong"); NAFFINE, *supra* note 16, at 99–117, 139–42 (generally discussing sanctity of personhood and relative worth of animals). See *supra* notes 32–37 and accompanying text for a discussion of Dworkin's humanist argument for viewing human life as intrinsically valuable. For humanists, this position is based on (or is at least consistent with) the well-known Kantian categorical imperative: "Act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means." IMMANUEL KANT, GROUNDING FOR THE METAPHYSICS OF MORALS 36 (James W. Ellington trans., Hackett Publ'g Co. 3d ed. 1993) (1785).

51. *See* GLOVER, *supra* note 50, at 50–51 (arguing that like racism, "speciesism"—the belief that human life has priority over animal life simply because it is human—is objectionable because it is morally arbitrary); TORCHIA, *supra* note 16, at 273 ("From a postmodern perspective, any attempt to impart a special dignity to humans (by virtue of an immaterial mind or soul, the basis of God's image in us) is guilty of 'specieism,' the unwarranted elevation of the human species over other species in the name of rational superiority or a certain sacrosanctness which renders human rights preeminent over the rights of other living things.").

52. *See* Will Kymlicka, *Liberal Egalitarianism and Civil Republicanism*: *Friends or Enemies?, in* DEBATING DEMOCRACY'S DISCONTENT: ESSAYS ON AMERICAN POLITICS, LAW, AND PUBLIC PHILOSOPHY 133 (Anita L. Allen & Milton C. Regan, Jr. eds., 1998) ("The [liberal] state does not justify its actions by reference to some public ranking of the intrinsic worth of different . . . ways of life, for there is no public ranking to refer to."); *see, also* KYMLICKA, *supra* note 8, at 212–19 (discussing liberal neutrality or "liberal egalitarianism" generally). Any objection based on neutrality raises the issue of what "neutrality" means. For example, John Rawls argues that a legal ban on first trimester abortions based on religious grounds would be an "unreasonable" imposition of a "comprehensive doctrine" in the political context. RAWLS, *supra* note 21, at 243 n.32; *see also id.* at 151–54 (arguing that reliance on religion in political decisionmaking must be carefully circumscribed). For an analysis of theistic conceptions of personhood, see generally NAFFINE, *supra* note 16, at 109–17. Naffine notes that, although a view based on a unique human sanctity can buttress important human rights,

> a dogmatic belief in human sanctity can have less beneficent, more illiberal effects. It lies behind attempts to impose severe constraints on human choice about some of the most personal aspects of existence, at some of the more vulnerable human moments: in pregnancy and at times of great suffering and in dying. Necessarily this places great strain on what is often said to be the first principle of our Rationalist liberal law, which is respect for individual autonomy and personal choice . . . .

Id. at 116. Naffine also objects that views based on human sanctity "can also countenance cruelty to animals . . . . [because in] the traditional and orthodox Christian view, animals are put on earth for our use; they do not have ends of their own that count." *Id.*

53. *See* GLOVER, *supra* note 50, at 42–45, 138–42 (critiquing "human life is sacred" argument).

fear if humans with disabilities are treated as no better than animals.[54] Such side effects arguments are also subject to the charge of speciesism because the differential treatment of higher mammals is justified solely by an arguably arbitrary distinction between unpleasant emotional effects on adult humans vis-à-vis the pain and psychic trauma of adult animals.

This Article is not about the rights that should be granted to animals as we know them today. The charge that the differential treatment of humans with disabilities vis-à-vis animals is speciesist has been raised herein solely to emphasize that the claim of humans' entitlement to the normative status as persons involves two components: (1) *all* humans are entitled to at least some degree of personhood in terms of such things as life, liberty, emotional well-being, and material prosperity; and (2) *only* humans are entitled to any meaningful degree of personhood. The two components are not necessarily inconsistent. Though the first is liberal, egalitarian, and inclusive, the second is not illiberal, inegalitarian, and exclusive *unless* humans deny personhood to an entity that has capacities equivalent to those human capacities, like complex reasoning and selfhood, that humans use to justify their claim to personhood. Absent some strong justification, 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.

Given that this potential inequality runs counter to the liberal theory of personhood, I argue that a manmade entity, like a machine with artificial intelligence, should be granted the basic Lockean right of self-ownership if it satisfies an appropriate test of the capacities required for autonomous personhood. To the extent an entity satisfies this test, the entity has a prima facie right to personhood—i.e., it should be accorded the status of a legally self-owning, autonomous person unless there is a very good independent reason to deny personhood.[55]

This argument is compatible with arguments that animals or artificial entities with a degree of self-consciousness that is very high, even though less than that of humans, should be granted some version of a lesser degree of personhood in the way that incompetent humans are.[56] However, there is no need to resolve herein the debate about granting such "animal rights" because this Article focuses on artificial entities with a level of intelligence and communication skills that are at least equivalent to that of an ordinary, adult human without any substantial disabilities. In contrast, higher mammals are, at most, entitled to a lesser level of personhood, which could be termed "sentient

54. *See id*. at 150–69 (discussing infanticide); *see also* Martha C. Nussbaum, *Animal Rights: The Need for a Theoretical Basis*, 114 HARV. L. REV. 1506, 1511 (2001) (reviewing WISE, *supra* note 48) (noting that special status for humans is important and that equating human and animal suffering could cause us to "lose our moral footing").

55. For discussion of loss of dominant species status as a possible reason for preventing development of or denying personhood to an entity with the requisite capacity, see *infra* Part V as well as *infra* notes 58–59, 115–18 and accompanying text.

56. *See* MARY MIDGLEY, UTOPIAS, DOLPHINS AND COMPUTERS: PROBLEMS OF PHILOSOPHICAL PLUMBING 115–17 (1996) (arguing that because higher apes are "highly sensitive social beings," treating them simply as things, rather than unique beings, "shocks morality"). For further discussion of the normative treatment of beings with lesser consciousness and intelligence than humans, see *supra* notes 38–52 and accompanying text.

being" and would resemble the status of a human child or mentally incompetent adult. The reason for denying any higher status to such animals is that they lack our self-consciousness and our ability to engage in complex thought and communication.[57] This reason is important herein because the capacities we use to distinguish ourselves from animals are the attributes to be used in the proposed test to determine whether an artifact, like the imaginary computer system, has the capacity to be a person.

The dispute about animal rights also differs from the concern herein because, unlike artificial entities like AIs, granting animals a lesser degree of personhood will not threaten the dominant status of the human species. In contrast, as many science fiction stories have indicated, a highly evolved computer "self" with the ability to utilize machines and weapons could be a serious candidate to replace humans as the dominant "species."[58] This risk of replacement provides, at the very least, a substantial prudential reason to consider whether and how to prevent such an entity from developing the physical and intellectual capacity for personhood and, if it does develop this capacity, from acquiring and exercising the right to autonomous personhood. To the extent we grant normative personhood to an artificial entity, we may want to be sure we retain, at least, our right to *equal* personhood. Moreover, if the entity has substantial competitive advantages, it might be prudent to deny or limit personhood for an artificial entity with such superior capacities that it could supplant humans as the dominant species even if we have the right to compete with it under conditions of equal personhood.[59]

III. ARTIFACTS AND PERSONHOOD

The modern rationalist claim to human uniqueness in terms of a right to personhood is based largely on the human capability for complex thought, communication, and technology. As indicated above, this claim of exceptional capability has been challenged in terms of animal rights. However, we have consistently rejected notions of the right of any other species on earth, regardless of whether it has some lesser degree of human capabilities, to personhood. Our reaction might be different if we encounter an extraterrestrial species—with a technology indicating that the species is at least our equal in terms of being "noble in reason" and able to exercise dominion over its environment—that demands treatment as autonomous persons. A challenge to the claim of uniqueness could also arise if an artificial entity claimed to have (or simply exhibited) both a highly developed capability of interacting with the world and a high level of intelligence and consciousness. Given advances in human technology,[60] this challenge appears more likely to occur in the near future than a visit from technologically advanced extraterrestrial aliens, which appears (at least, to most of us) not to have happened at

57. See *supra* notes 38–43 and accompanying text and *infra* Parts III.A.1–2 for the importance of intellectual skills and self-consciousness to personhood.

58. See *infra* notes 305–07, 347–64 and accompanying text for discussion of science fiction treatments of this scenario.

59. See *infra* Part V as well as *infra* notes 115–18 and accompanying text for a more complete discussion of the dominant species problem.

60. *See infra* notes 146–69, 196–203, and accompanying text.

any time in human history.[61] The response to this challenge will require a test for determining whether the artifact has the capabilities required for autonomous personhood. The response will also require a consideration of whether, if those capabilities are present, we should grant it the legal status of a person or of a thing that can be owned. In considering both the capacity for and the right to personhood, this Article focuses solely on the legal issue of self-ownership. Though important, more specific legal rights are beyond the scope of the analysis herein.

A. Test of Capacity for Personhood

Regardless of the nature of a challenge to human uniqueness, addressing the challenge will require a standard of the attributes and capabilities that a challenger must satisfy in order to be the equivalent of a human in terms of personhood. An entity which passes the test would be regarded as a conscious being like, but not the same as, a human. The standard for the *capacity* for personhood proposed herein is a behavioral test. It requires that an entity exhibit behavior demonstrating: (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. An entity that passes this test is, unlike animals, entitled to at least a prima facie right to be treated as a person rather than property.

1. Complex Intellectual Interaction

In order to be a "living" entity of any sort, one must have the ability to interact meaningfully with the environment by receiving and decoding inputs from, and sending intelligible data to, its environment.[62] The entity must be both rational and capable of learning from its experiences in these activities.[63] The ability to interact with the environment is a minimal requirement that can be satisfied by animals and some existing machines. The claim of a human right to personhood (and of a lack of a right for animals) is based on more complex skills, particularly our ability to engage in complex thought and communication. Thus, the entity's communication with its

61. For an argument that we are very unlikely to encounter another technologically advanced species anytime soon, see RAY KURZWEIL, THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY 342–49, 357–59 (2005).

62. *See* MARY MIDGLEY, THE ETHICAL PRIMATE: HUMANS, FREEDOM AND MORALITY 10 (1994) (questioning whether "life without sense-perception . . . could intelligibly be called life at all"); SIDNEY PERKOWITZ, DIGITAL PEOPLE: FROM BIONIC HUMANS TO ANDROIDS 147–72 (2004) (discussing robots' need for senses of vision, hearing, touch, taste, and smell). The requirements should not be too stringent in terms of a list of necessary senses. Instead, as with humans, who sometimes lack senses like sight or hearing, the emphasis should be on functional interaction rather than on a particular mode of interaction.

63. *See, e.g.*, J. STORRS HALL, BEYOND AI: CREATING THE CONSCIENCE OF THE MACHINE 116 (2007) (discussing importance of learning and growth and asking: "Will we decide a program is a true AI if it can pass the Turing Test—in the sense of having a broad set of capabilities, memories, and knowledge comparable to an adult human—but cannot learn and grow?"). The Turing test is discussed *infra* at notes 170–86 and accompanying text. *See also* Daniel Dennett, *Conditions of Personhood*, *in* THE IDENTITIES OF PERSONS 175, 177 (Amélie Oksenberg Rorty ed., 1976) (stating that moral "persons are rational beings" (emphasis omitted)).

environment must be sufficiently diverse and sophisticated that we can view it as the product of complex thought.[64]

In all likelihood, most candidates for personhood will be able to interact physically with the world. For example, a computer entity could interact with the world through remotely controlled robotic machines. Although the ability to interact *physically* with the environment is often viewed as necessary for life and personhood,[65] it is theoretically possible for an entity, with only the ability to analyze and communicate with its environment, to "live" solely in the form of data.[66] This entity could take many forms, including a copy of a human that can be transferred to and "live" totally in a virtual world within a computer. Because the nature of any such copying is largely speculative today,[67] it will not be addressed herein except for a brief listing of possible issues involved.[68] This Article also contains a short discussion of science fiction treatments of the concept, including digital entities limited to data interconnection and digital entities able to interact physically using remote control of robots.[69]

2. Self-Consciousness

We distinguish ourselves from animals not only in terms of the complexity of our thought and communications, but also in terms of our sense of being individual selves. Thus, an essential aspect of personhood is having a sense of being a "self" that not only exists as a distinct identifiable entity over time but also is subject to creative self-definition in terms of a "life plan"—i.e., a plan for living a unique life story over a relatively substantial period of time.[70] Although we all share this sense of self-

64. *See* Dennett, *supra* note 63, at 178 (arguing that necessary conditions for *moral* personhood include being "capable of verbal communication" (emphasis omitted)).

65. *See* MIDGLEY, *supra* note 62, at 10 (questioning whether "life consisting solely in the arrangement of abstract 'information' . . . could intelligibly be called life at all"). Where robots are concerned, physical interaction is necessary; *see also* PERKOWITZ, *supra* note 62, at 123–45 (discussing robots' need for limbs, hands, mobility, and expression).

66. *See* KURZWEIL, *supra* note 61, at 260 (disagreeing with need for physical embodiment to affect world and arguing central concern should be intelligence).

67. For an argument that it will be possible to upload (or download) ourselves into a machine in the near future, see KURZWEIL, *supra* note 61, at 200 ("[T]he end of the 2030s is a conservative projection for successful uploading."). Rodney Brooks, a former director of the MIT Artificial Intelligence Laboratory, notes that because this process will be incredibly complex and difficult, "[w]e are a long, long way from being able to download ourselves into computers or robots." BROOKS, *supra* note 38, at 206, 208. He also notes that persons, including Kurzweil, making optimistic predictions over the past decades tend to predict success at the point in time when they will be seventy-years-old, and thus in need of the technology. *Id.* at 206. For further discussion on this topic, see KAKU, *supra* note 49, at 53–60, 87–95, 113–15 (discussing brain-imaging technology, computer modeling of the brain, downloading of human brains to robotic bodies, and issue of whether the benefits of downloading would be worth the costs).

68. See text accompanying and following *infra* note 164 for a discussion of issues that may arise if the uploading of human consciousness into machines ever became possible.

69. See *infra* notes 274–75, 375–76, and accompanying text.

70. See *supra* notes 17–37 and accompanying text for a discussion of Rawls's concept of a life plan and Dworkin's concept of human responsibility "to create a life." *See also* DAVID DEGRAZIA, HUMAN IDENTITY AND BIOETHICS 6–7, 77–114 (2005) (defining person as "someone with the capacity for complex forms of consciousness," analyzing concept of "self-creation," and discussing "narrative identity" as a narrative of one's

consciousness, we are not sure about its nature or origin.[71] Nor are we sure about the nature and role of our minds.[72] What exactly is "the juice" that makes us so special?[73] The behavioral standard adopted herein sidesteps the issue of "the juice" by focusing on behavior that indicates self-consciousness, rather than on metaphysical questions concerning the nature of our self-consciousness.

Defining one's self as a unique individual over time requires some degree of imagination or "creativity" in designing and implementing a life plan.[74] Artificial entities need not be creative geniuses in order to satisfy this standard; they only need to have dreams or visions of what they want to be and do in their lives and an understanding of how to plan and effectuate these dreams. Humans vary enormously in their capacities for creativity in envisioning the future. None of us is creative all the time; habit and routine are parts of everyone's life. Similarly, the entity does not have to think and communicate at the level of a genius. Ordinary human levels of complexity are sufficient.

Because feelings and emotions are so central to choices a human makes in terms of a life plan, it is logical to consider their role in the requirements for personhood. To the extent that only basic emotions are required, modified humans and modified animals would probably satisfy the requirement because humans and animals exhibit emotional behavior. Machine entities are more problematic. The robotic machines that exist today have goals, but we have no reason to believe they "care" whether the goals are satisfied.[75] By and large, machines today, no matter how complex, are simply robots with the ability to process data quickly enough to give the appearance of emotion. Despite this apparent emotional dimension, these machines are, in effect, complex versions of a thermostat, which is simply a machine with a goal or purpose of seeking an equilibrium temperature even though it lacks "concern" about achieving that goal.[76] In order for a machine to go beyond being like a thermostat and become a self-

life (emphasis omitted)); PERKOWITZ, *supra* note 62, at 173–97 (discussing robotic thinking, emotions, and self-awareness); RAWLS, *supra* note 18, at 561 (arguing that "the unity of the person is manifest in the coherence of his plan" for life); Dennett, *supra* note 63, at 178 ("[T]here is a way in which *we* are conscious in which no other species is conscious. Sometimes this is identified as *self*-consciousness . . . ."); *cf.* BROOKS, *supra* note 38, at 103 (arguing that intelligent entities must be "situated" in the sense of having "an ongoing existence that is tied into the flow of time").

71. *See* DANIEL C. DENNETT, BRAINCHILDREN: ESSAYS ON DESIGNING MINDS 131–39 (1998) (noting various theories of consciousness and lack of consensus on topic).

72. See *infra* note 191 and accompanying text for differing views on the role of the mind.

73. The term and concept of "the juice" is taken from BROOKS, *supra* note 38, at 187–88.

74. *See* DEGRAZIA, *supra* note 70, at 77–114 (discussing "self-narrati[on]" and "self-creation"). See *supra* notes 3–4 and accompanying text for a discussion of Pico's view of humans as "the free and proud shaper of . . . [their] own being[s]." See also *supra* notes 27–37 and accompanying text for a discussion of Dworkin's view that a person "creates his life."

75. *See* BROOKS, *supra* note 38, at 157 (asserting that placing an "emotional system" in a robotic toy is not the same as the toy having "real emotions"); PERKOWITZ, *supra* note 62, at 182 (noting consensus among robotic and computer experts that artificial beings do not possess "intrapersonal intelligence").

76. *See* W. ROSS ASHBY, AN INTRODUCTION TO CYBERNETICS 81, 222 (1956) (discussing machines, like thermostats, as goal-seeking regulators designed to be stable around a state of equilibrium); DENNETT, *supra* note 71, at 327 (claiming that a thermostat can be viewed as a system that "has a rudimentary goal or desire (which is set, dictatorially, by the thermostat's owner, of course), which it acts on appropriately whenever it believes (thanks to a sensor of one sort or another) that its desire is unfulfilled").

conscious entity with a life plan, the machine must somehow care about the success of the plan.[77] Such caring requires, at the very least, two emotional concerns. First, the entity must care about its survival.[78] Second, it must feel there is a purpose or reason beyond mere survival for its life. In order to develop a life plan, an entity must have a sense of what gives its life "meaning."[79]

The requirement of a meaning in life could be fairly minimal. For example, it could be sufficient for the entity to exhibit what Rawls refers to as the "Aristotelian Principle," which he defines as follows: "[O]ther things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity."[80] Rawls asserts that this principle plays a central role in a person's judgments of value and choice of a rational long-term plan.[81] As a result, if a person, who supports himself "by solving difficult mathematical problems for a fee," enjoys the complexity of counting blades of grass in geometrically shaped lawns, "a rational plan for him will center around this activity."[82] If the entity taking pleasure in this way is a machine, rather than a human, this approach to a life plan should be sufficient to satisfy the requirement of feelings required for autonomous personhood.

Imposing a stronger, more complex set of goals or emotions is very questionable. Some feelings—for example, emotions like rage and jealousy—often undermine our ability to be "noble in reason." Other strong feelings that serve Darwinian needs for us might not be necessary for rational machines. For example, a machine intelligence would need programming to address actual or potential harm in a rational manner,[83] but would this have to be pain as we know it? Similarly, though the entity described by Rawls might experience some negative feeling akin to sadness if it cannot count blades of grass, would deep sorrow, despondency, or depression be necessary to personhood? Imposing such additional requirements in terms of feelings would be simply arbitrary and speciesist if the requirements are defended solely on the assertion that humans have

77. *See* DENNETT, *supra* note 71, at 153, 164, 169 (designers of a "conscious robot" must include a "motivational structure" so that success in achieving its preferences will "matter" to it (internal quotation marks omitted)); MIDGLEY, *supra* note 56, at 168–73 (arguing that "[c]onsciousness is the condition of active beings, the condition of acting and suffering and enjoying"); MIDGLEY, *supra* note 62, at 10, 17 (questioning whether "a life without . . . emotion . . . consisting solely in the arrangement of abstract 'information' would be a human life, or indeed anything that could intelligibly be called life at all" and arguing that humans are both egoists and altruists).

78. *See* DENNETT, *supra* note 71, at 164 (stating that conscious robot's "motivational" structure must include strong desire not "to engage in self-destructive activity").

79. *See* Steven D. Smith, *Believing Persons, Personal Believings: The Neglected Center of the First Amendment*, 2002 U. ILL. L. REV. 1233, 1274 (2002) (stressing, in context of theory of First Amendment, concept of person as a being with beliefs, particularly "beliefs concerned with what is often described as 'meaning,' or perhaps 'ultimate meaning' or 'ultimate purpose'—beliefs that address, as we sometimes say, 'the point of it all'").

80. RAWLS, *supra* note 18, at 374.

81. *Id.* at 374–80.

82. *Id.* at 379–80.

83. *See, e.g.*, HANS MORAVEC, ROBOT: MERE MACHINE TO TRANSCENDENT MIND 123 (1999) (discussing a "watchdog program" that would pull a robot away from harmful experiences and negatively condition encountering them again).

these characteristics or that we simply cannot conceive of self-consciousness that lacks such richer emotions. Moreover, requiring a wide range of emotions is inconsistent with our recognition of humans as persons even where they act as if they lack a full range of human emotions.[84] Programming a machine to have emotions akin to extreme feelings that are not necessary to survival and achievement of goals, but are programmed into the machine in order to make it more like humans, would also raise serious ethical issues concerning our obligations to the sentient artifacts we create in this way.[85]

3. Community

The claim of a right to personhood only matters within a community of autonomous persons. A human in isolation has little reason to worry about the treatment of or by other people. In contrast, a person who exists in a community with other persons must be able to interact responsibly as a member of that community. (This interaction will also have a strong impact on how a self-conscious entity defines the nature and content of its selfhood.)[86] A person's claim of a right to personhood

84. *See infra* notes 179–83 and accompanying text for a discussion of circumstances in which humans, despite various states of emotional impairment, are still regarded as persons.

85. In 1978, Daniel C. Dennett published a book which expressed skepticism about "artificial or synthetic pain" that would be analogous to artificial intelligence because of: (1) the "irredeemable incoherency in our ordinary concept of pain," and (2) concerns that pain was bound up with a number of basic conceptions of humans, including "our ethical intuitions, our senses of suffering, obligation, and evil." DANIEL C. DENNETT, BRAINSTORMS: PHILOSOPHICAL ESSAYS ON MIND AND PSYCHOLOGY 197, 228 (1978). In particular, he expressed a concern about whether "any attempt at robot synthesis of pain can be conducted independently of questions about what our moral obligations to this robot might be." *Id.* at 197–98. In a set of essays published in 1998, Dennett discussed a project to construct a "conscious robot" called "Cog" and argued that such a robot must have a motivational structure akin to a human. DENNETT, *supra* note 71, at 163–64. In order to create a robot with such a structure,

> Cog's creators [decided] to make Cog as much as possible responsible for its own welfare. Cog will be equipped with some innate but not at all arbitrary preferences, and hence provided of necessity with the concomitant capacity to be "bothered" by the thwarting of those preferences, and "pleased" by the furthering of the ends it was innately designed to seek. . . . Cog may be said to have quite crude, simplistic, one-dimensional pleasure and pain, cartoon pleasure and pain if you like, but then the same might also be said of the pleasure and pain of simpler organisms—clams or houseflies, for instance. . . . The reasons for saying that something *does* matter to Cog are not arbitrary; they are exactly parallel to the reasons we give for saying that things matter to us and to other creatures. . . . [I]t will be interesting to see if the skeptics have any good reasons for declaring Cog's pains and pleasures not to matter—at least to it, and for that very reason, to us as well. It will come as no surprise, I hope, that more than a few participants in the Cog project are already musing about what obligations they might come to have to Cog, over and above their obligations to the Cog team.

Id. at 169. Others have expressed similar views. According to Brooks,

> those [machines] that we make more intelligent, that we give emotions to, and that we empathize with, will be a problem. We had better be careful just what we build, because we might end up liking them, and then we will be morally responsible for their well-being. Sort of like children.

BROOKS, *supra* note 38, at 195.

86. *See* KYMLICKA, *supra* note 8, at 221–28 (discussing whether it is possible to be a self that is not encumbered by being embedded in existing social practices). See also *supra* notes 47–48 and accompanying text and *infra* notes 102–03 and accompanying text for a discussion of the concept of a child as a "person in training."

presupposes a reciprocal relationship with other persons who would recognize that right and who would also claim that right—i.e., persons in the community must recognize (either on the basis of respect or prudential self-interest) each other's personhood and must be held responsible for violations of that right. Addressing these reciprocal claims raises questions concerning the meaning of community and the attitude or viewpoint towards the community that an artificial entity must possess in order to be a member of a community of persons. In large part, the questions are interconnected because the answers to both depend upon whether recognition of personhood must be based on a sense of personal respect or simply on prudential self-interest.

Rawls's distinctions concerning three types of communities provide a helpful framework for addressing these questions. The first type is a closely knit community sharing a "comprehensive philosophical doctrine" concerning basic personal, religious, and political values.[87] The second type of community arises in modern democratic societies, which are characterized "by a pluralism of incompatible yet reasonable comprehensive doctrines," which have areas of overlap that provide a shared consensus on basic political values.[88] This community is defined in terms of Rawls's distinction between being *rational*—viewed as the capacity to use judgment and deliberation to seek one's individual ends and to balance competing ends in a plan of life—and being *reasonable*, which involves a readiness to accept and cooperate with "a system of fair cooperation and . . . fair terms . . . [that is] reasonable for all to accept."[89] The concepts of rational and reasonable are complementary and "work in tandem to specify the idea of fair terms of cooperation."[90] Both are necessary to the second type of community because "[m]erely reasonable agents would have no ends of their own they wanted to advance by fair cooperation; merely rational agents lack a sense of justice and fail to recognize the independent validity of the claims of others."[91]

Rawls used the above distinctions to define the third type of community as one in which the actors are rational, but not reasonable. He describes this third category of community as one based on a "modus vivendi" (manner of living), which is a term used to describe political and economic arrangements based on an acceptance of terms and conditions that are sufficiently beneficial to both opposing parties that neither sees a challenge to the status quo as advantageous.[92] Because an arrangement where only some parties have rights is not likely to be beneficial to opposing parties lacking rights, a modus vivendi is more stable where the members of each side have rights vis-à-vis the other. The stability of such an arrangement of reciprocal rights and obligations "is contingent on circumstances remaining such as not to upset the fortunate convergence

87. RAWLS, *supra* note 21, at xvi.

88. *Id.*; *see also id.* at 133–72 (discussing overlapping consensus). Unreasonable comprehensive doctrines should be contained "so that they do not undermine the unity and justice of society." *Id.* at xvi–xvii.

89. *Id*. at 48–51.

90. *Id.* at 52.

91. *Id.* For a more complete development of the role of persons' treatment of the moral claims of others, see STEPHEN DARWALL, THE SECOND-PERSON STANDPOINT: MORALITY, RESPECT, AND ACCOUNTABILITY 91–118 (2006).

92. *See* RAWLS, *supra* note 21, at 147 (discussing concept of "modus vivendi" in terms of social and political cooperation).

of interests."[93] Such a contingent arrangement among rational egoists is very different from the second type of social and political community, which is composed "of citizens as reasonable and rational, and free and equal."[94]

Rawls's distinction between being rational and being reasonable parallels H.L.A. Hart's distinction between being obligated and having an obligation. Where members of a society generally accept a scheme of legal rules, the rules are obeyed partly on the basis of an internal sense of obligation to obey the law.[95] In contrast, where obedience to the system's rules is based primarily on fear of punishment by a legal system analogous to a regime of gunmen, people only feel obliged to obey and would be willing to disobey in situations where punishment can be evaded.[96] To Hart, it is possible to have a legal system where citizens obey laws because they are obliged to do so, but "in a healthy society they will in fact often accept these rules as common standards of behavior and acknowledge an obligation to obey them."[97]

This Article adopts Rawls's third category of community, based on a modus vivendi, as the test for autonomous personhood for three reasons. First, humans are likely to resist the idea that machines could have a sense of justice and fair cooperation and the idea that machines are entitled to be treated justly. This resistance is understandable because of the gap between the nature and characteristics of biological and mechanical identities.[98] There is also a substantial possibility that humans would simply refuse to accept the concept of a machine with a moral sense no matter how strong the evidence was to support a finding that a machine's behavior exhibited that sense.

Second, human political and legal communities do not necessarily require anything more than a modus vivendi. Communities based on shared views about justice

93. *Id.*; *see also* JOHN RAWLS, THE LAW OF PEOPLES WITH "THE IDEA OF PUBLIC REASON REVISITED" 149 (1999) (arguing that Catholics and Protestants honored the principle of toleration only as a modus vivendi in the sixteenth and seventeenth centuries).

94. RAWLS, *supra* note 21, at 149. Mary Midgley criticizes such arrangements on the basis of the following arguments concerning the qualities of a shared community vis-à-vis a modus vivendi and the difficulty of imagining how a modus vivendi among rational egoists "would count as a culture at all":

> Mere mutual terror, driving coexisting egoists to form a social contract, could certainly not produce one. It would never generate the myriad positive activities that go to make up a culture. The common standards, common ideals, common tastes, common priorities that go to build a common morality rest on shared joys and sorrows, and all require active sympathy. Morality needs, not just conflicts, but also the willingness and capacity to look for shared solutions to them. Morality, as much as language, seems to be something that could only occur among naturally sociable beings.

MIDGLEY, *supra* note 56, at 144. Midgley's argument has validity in terms of humans; as "naturally sociable beings" we need richer connections like those in the first and second types of communities in Rawls's scheme. However, in terms of political and legal communities, modus vivendi is sufficient, though not ideal, for accommodating more closely knit subcommunities of "naturally sociable" humans within a larger political community. For another type of response to Midgley, see *infra* note 99 and accompanying text.

95. H.L.A. HART, THE CONCEPT OF LAW 56–58, 82–91 (2d ed. 1994).

96. *Id.* at 82–85.

97. *Id.* at 116. Hart also argues that, in order to distinguish the rule of law from a system of rule by gunmen, legal officials should recognize a sense of obligation to obey rules that are valid according to "common standards of official behavior." *Id.* at 116–17.

98. See *infra* notes 174–95 and accompanying text for a discussion of relevant differences between biological and mechanical identities.

and morality are more desirable in terms of justice and stability. However, both Rawls and Hart recognize that political and legal systems can exist solely on the basis of rationality and prudence as reasons to obey rules. Moreover, Rawls's preference for the overlapping consensus of the second type of community has been criticized as too stringent for political arrangements in an increasingly pluralistic world, where a modus vivendi is the best we can hope for in many political systems.[99] This model of rational self-interest is also central to "[t]he task of economics . . . [which] is to explore the implications of assuming that man is a rational maximizer of his ends in life, his satisfactions—what we shall call his 'self-interest.'"[100] Similarly, international law provides a model where artificial actors, "persons" in the form of nation-states, recognize reciprocal rights and responsibilities on the basis of a modus vivendi.[101] Such rational, self-interested behavior may characterize initial relationships between humans and artificial entities. For example, humans may be reliant on a self-conscious computer handling vital functions and may be rationally concerned that it can damage human interests. If there is a corresponding concern by the machine for its own self-interest, a modus vivendi, based on a rational assessment of costs and benefits, would benefit both sides.

Finally, imposing a stronger measure of community as a requirement for personhood would present a classic nature/nurture conundrum: Can we create an entity with the capacity for personhood in terms of being a reasonable actor (in Rawlsian terms) without treating it as a person as it develops its capacities? Like human children, it may be necessary to treat intelligent machines as "persons in training."[102] As John Stuart Mill noted, granting humans freedom in thought and expression is necessary to enable both great thinkers and "average human beings to attain the mental stature which they are capable of."[103] Similarly, in order for artificial entities to be part of a Rawlsian pluralist political community with an overlapping consensus concerning the basic political framework, it may be necessary to grant them personhood initially on the basis of a modus vivendi.

99. *See* JOHN GRAY, TWO FACES OF LIBERALISM 1, 23 (New Press 2000) (arguing that Rawls's scheme of shared consensus in the second type of community is based on a narrow "Americocentric version" of political society and that more emphasis should be placed on the task of "refashioning liberal toleration so that it can guide the pursuit of *modus vivendi* in a more plural world").

100. RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 3 (7th ed. 2007) (footnote omitted). Posner notes that because self-interest can include concern for the happiness or misery of others, it is preferable "to speak of 'utility' rather than of self-interest" and a "[c]entral" assumption of economic analysis is that "man is a rational utility maximizer in *all* areas of life." *Id.* at 4. Rawls himself notes that "rational agents . . . [are not necessarily] solely self-interested: that is, their interests are not always interests in benefits to themselves." RAWLS, *supra* note 21, at 51.

101. RAWLS, *supra* note 93, at 27–29. Even two neighboring societies at war with one another can develop a modus vivendi system to govern cross-border interaction. *See generally* Peter T. Leeson, *The Laws of Lawlessness*, 38 J. LEGAL STUD. 471 (2009) (examining use of novel institutions and cross-border criminal law to resolve sixteenth century Anglo-Scottish border conflicts).

102. See *supra* notes 47–48 and accompanying text for a discussion of human infants as "persons in training." For science fiction treatment of the importance of accepting and nurturing our creations, see *infra* notes 262–64.

103. MILL, *supra* note 18, at 974.

A modus vivendi requires that an entity have a sufficient ability to understand the human point of view concerning rights and duties within reciprocal self-interested relationships.[104] In contrast to an entity's own life plan, it is not necessary that the entity care emotionally about humans and their concerns. Understanding a human's point of view and acting rationally in response to that perspective is sufficient. Modified humans and animals will probably have this capacity. Humans can empathize, and higher order mammals like dogs have some ability to understand our feelings. Machines also have some degree of this ability because, from a behavioral perspective, they can be (and often have been) programmed[105] to respond in words and actions in a way that humans interpret as showing "concern" about a human's situation.[106]

> For a robot to develop empathy for a human being, it seems likely that the robot will need to observe that person's behavior in different situations, then make intelligent guesses as to what is going on in that person's mind in a given situation, in order to predict subsequent behavior. The acquisition of empathy is therefore essentially a learning task—relatively easy to implement in robots.[107]

Currently, such displays of empathy are properly viewed as simulated emotions that are not related to feelings experienced by the robot.[108] However, even simulated emotions are sufficient so long as the machine's behavior is responsive to the humans' reasons for engaging in the modus vivendi. Moreover, in time, as machines consistently display a sense of empathy in terms of reciprocal recognition of rights and obligations, humans may even come to accept the idea that a machine, like a human, can act in accordance with the moral requirements of not only the second, but also the first type of Rawlsian community. In time, we may even think they have feelings like us. We might even be right, but will that matter?

104. See *supra* notes 92–94 and accompanying text for a discussion of modus Vivendi. *Cf.* DARWALL, *supra* note 91, 4–27 (discussing importance in morality of showing concern for claims of others (second person actors) and noting distinct role first person actor's relationship with second person plays in terms of reasons for first person's actions and second person's claims); Dennett, *supra* note 63, at 178 (discussing view that a moral person must be capable of reciprocating with other persons).

105. *See* BROOKS, *supra* note 38, at 64–65, 91–97 (discussing ability of robotic head with limited speech abilities ("nonsense syllables") to engage humans by responding behaviorally as if conversation is occurring); PERKOWITZ, *supra* note 62, at 181–84 (discussing ability of robots to interact responsively to human behavior); Cynthia Breazeal, *Human-Robot Partnership*, IEEE INTELLIGENT SYS., July-Aug. 2006, at 80 (discussing ability of robots to use "social-cognitive skills to compare and reason about . . . its human partner's goals and belief states").

106. *See* DENNETT, *supra* note 71, at 169, 340 (discussing how researchers, who are aware that a robot lacks consciousness, tend to treat it as a conscious being); LEVY, *supra* note 39, at 78–79 (discussing human treatment of robots as if machines were gendered), 97–103 (discussing human response to robotic dog as if it were a dog).

107. LEVY, *supra* note 39, at 107.

108. See *supra* notes 75–76 and accompanying text for a discussion of current emotional capacities of robots.

4. Application

The proposed standard of capacity for personhood relies on our ability to analyze an entity's behavior and determine whether it exhibits complex intellectual interaction, a sense of self, and an ability to be a member of at least a modus vivendi community. These tasks will be difficult because the test is abstract and vague and will require us to draw potentially arbitrary lines in matters, like levels of ability to engage in complex thought, that differ in degree rather than kind. In addition, the standard does not address the details involved in difficult issues about mind and consciousness that arguably play an important role in defining personhood.[109] As a result of these problems, judgment and normative perspectives will become involved in evaluating behavior. In particular, there will be a risk that application will be affected by possible human bias and concern for loss of human uniqueness. Though daunting, these problems should not be overemphasized. We manage to address issues of human personhood even though issues about the nature of the human mind and about human self-consciousness and identity are far from solved.[110] We also continue to struggle in determining which of us is entitled to autonomous personhood. For example, there remain legal disputes concerning involuntary commitment and cultural disagreement concerning women's rights. Despite all its difficulties, the standard herein is sufficient as an exploratory account of how to assess whether artifacts should ever be entitled to self-ownership. Broader issues of what personhood entails in terms of civil and political rights can be addressed after this more basic question is addressed.[111]

B. Legal Personhood When Capacity Test Is Met

While application of a test for determining capacity for personhood is meant to be an empirical matter based on behavior, the decision to grant or recognize the right of autonomy as a legal person, whether for a human or an artifact, is a matter of law and politics. As a result, granting the legal status is far less "objective" than decisions concerning capabilities. For example, adult women without handicaps clearly have the capabilities required for personhood, but are nonetheless still denied aspects of that legal status in many cultures. There are many reasons that might be given for such denials of autonomous personhood to some humans, including normative arguments based on different cultural frameworks than western liberalism and selfish prudential concerns like power and economic advantage disguised as claims of natural right or superiority. From the perspective of egalitarian liberal personhood, however, such

109. This point will be addressed more fully in the discussion of machine intelligence at *infra* notes 191–94, 204–13, and accompanying text.

110. See *infra* notes 191, 204–06 and accompanying text for a discussion of disagreement concerning human mind, self-consciousness, and identity.

111. See *infra* Part III.B.3, *infra* note 157 and accompanying text, text following *infra* note 164, as well as *infra* notes 209–13, 241–42, and accompanying text for issues concerning civil and political rights. For an interesting and useful discussion of these details, as well as a scenario for initial recognition of personhood for "artilects" (artificial intellects), see Frank W. Sudia, *A Jurisprudence of Artilects: Blueprint for a Synthetic Citizen*, KURZWEIL (Aug. 21, 2001), *available at* http://www.kurzweilai.net/meme/frame.html?main=/articles/art0270.html?m%3D4.

unequal treatment is wrong[112] and it is easy to criticize such reasons for unequal treatment as simply culturally biased or selfish. However, criticisms like these have been used in assessing our culture's treatment of animals,[113] and thus we cannot simply assume we are immune from cultural bias and selfishness. Instead, given our treatment of animals, the initial reaction to the possibility of personhood for an artifact is likely to be negative, particularly if an intelligent machine is involved. Therefore, the following analysis starts with the reasons for denying personhood to an entity with the capacity for personhood.

1. Reasons to Deny Personhood

One reason for denying personhood to humans, who clearly have the capacity for personhood, is that it is very useful to have a slave, whether human or mechanical, so long as the benefits exceed the costs. For many humans, the emotional benefits of owning and controlling a slave may make slavery desirable even if it is not economically efficient.[114] If one wants to have the feeling of being moral, it is possible to "justify" this position by a racist or speciesist argument that the human or entity enslaved is simply not entitled to the status of personhood. One can put reason aside and simply refuse to accept the validity of any assertion that the entity has the capacity for personhood—for example, by simply asserting that a machine or an enhanced animal can never be person. However, such dogmatic assertions in the face of contrary empirical data of the capacity for personhood are no more legitimate than the dogmatic claim that enslavement of Africans is permissible because they are "inferior" beings.

Another, more justifiable, reason for denying personhood to artifacts is the desire to reduce or eliminate a threat to the dominance of the human species.[115] This goal might be achieved by using the human slavery model to deny normative personhood to a class of entities with the capacity for personhood. There are problems, however, with the slavery approach. Apart from normative objections, this approach raises practical problems of implementation, particularly the problem of subjugating the entities capable of personhood. More complex issues could also arise, such as whether an owner of such an artificial entity would be allowed to free the slave and if so, what the status of the freed slave might be.[116] In order to avoid these problems, we might try to restrict technological development of artificial entities so that they cannot develop the physical and intellectual capability for personhood. However, as indicated below, this

112. See *supra* Part II.A and accompanying text for a discussion of human equality.

113. *See supra* notes 47–54 and accompanying text for discussion of criticisms of treatment of animals by humans.

114. *See, e.g.*, WITOLD RYBCZYNSKI, A CLEARING IN THE DISTANCE: FREDERICK LAW OLMSTEAD AND AMERICA IN THE NINETEENTH CENTURY 113–27 (1999) (summarizing Olmstead's travels in the antebellum south and his observations on the allegiance to slavery by owners, the economic inefficiencies of slavery, and slavery's corrupting influence on values of society). See also *infra* notes 330–32 and accompanying text for a discussion in the science fiction context of the "pleasure . . . of being taken care of by your own robots."

115. See *infra* Part V for an analysis of the dominant species problem.

116. *See generally* JOAN SLONCZEWSKI, DAUGHTER OF ELYSIUM (Avon Books 1993) (depicting successful "robot liberation" movement led by robot whose owner's open treatment of her as a person had been tolerated by society).

approach may be hard to implement.[117] A variation of this approach would be to exterminate or "lobotomize" artificial beings that, having achieved the capabilities of personhood, are a threat to humanity's status. As with limiting technological development at an earlier stage, this approach may have substantial practical problems. In addition, any form of extermination is, in effect, a form of genocide of a "species" of self-conscious persons. The normative and practical problems of denying personhood in order to protect human dominance will be discussed in more detail below.[118]

Limitations on the freedom of artificial entities can also be justified in terms of the Lockean approach to property. In Locke's scheme, each human has a right to the product of his labor.[119] People could rely on this right to assert ownership to artificial persons they produce. Once again, however, speciesism objections arise. The Lockean premise does not apply to children. Even though all adults owe not only their existence but also their food, clothing, and education as children to other humans, they incur no debt for these benefits. Instead, they are entitled to personhood simply because they are human—i.e., they are "creatures of the same species and rank."[120] Granting children personhood without legal obligations,[121] but not artificial persons, is not necessarily arbitrary. Parents understand that they are investing in persons, not things. In contrast, a corporation which invests millions in a supercomputer expects a return on that investment.[122] This problem could be addressed, however, by a scheme[123] that required an artificial person to pay back the cost of producing the artificial person.[124]

117. See *infra* Part V.A and accompanying text for a discussion of difficulties in restricting the development of cloning and machine-intelligence technologies.

118. See *infra* Part V.

119. See *supra* note 8 and accompanying text.

120. LOCKE, *supra* note 6, at 404.

121. Children may owe moral obligations to parents, but legal obligations are rare.

122. *See* DAVID GERROLD, WHEN HARLIE WAS ONE (RELEASE 2.0) 68–70 (rev. ed. 1988) (discussing problem of losing millions of dollars of investment in a self-conscious AI if it is granted personhood).

123. *See, e.g.*, IAIN M. BANKS, CONSIDER PHLEBAS 260 (1987) (depicting scenario where any self-conscious "Accredited Free Construct, certified sentient under the Free Will Acts" would be totally free once "Incurred Generation Debt" was completely paid); *cf.* GERROLD, *supra* note 122, at 83–88 (depicting motivation of sentient AI to provide corporation a return on investment because "[f]or him to use the company's equipment and electricity without producing something in return would be suicidal"). For further discussion of the fictional world depicted in *Consider Phlebas*, see *infra* notes 371–73 and accompanying text. An approach like this is also used in a series of books by Anne McCaffrey in which children with severe physical handicaps are encapsulated in a life-support shell and trained to use their brains to control a spaceship or other complex machine. After training, each of these cyborgs must pay off the cost of the training, medical procedures, and equipment that were required to achieve full cyborg capabilities. Except for the requirement of paying this debt, the cyborgs are regarded as persons, though many normal humans have trouble relating to them. *See* ANNE MCCAFFREY, THE SHIP WHO SANG (1969) (edition includes five stories by McCaffrey previously published as: DRAMATIC MISSION (1969); THE SHIP WHO DISSEMBLED (1969); THE SHIP WHO KILLED (1966); THE SHIP WHO MOURNED (1966); THE SHIP WHO SANG (1961)). See also *infra* note 273 for further discussion of McCaffrey's cyborgs. McCaffrey has also co-authored books in the series. *See, e.g.*, ANNE MCCAFFREY & S.M. STIRLING, THE CITY WHO FOUGHT (1993).

124. If this indentured servanthood approach is adopted, it should be viewed simply as a debt owed by a person. The person with the obligation to repay should not be viewed in terms of the status of a slave or of a servant with severely restricted rights. Nor should it be impossible to repay the cost of production. For example, if the actual costs of developing a machine intelligence are too large for repayment, the corporation should bear the cost in excess of what is affordable because spreading risks has long been one reason for

2. Reasons to Grant Personhood

We could decide to grant the status of legal personhood to artificial entities with the capacity for personhood on the basis of liberal equality. Locke argued for equal human rights, including the right of self-ownership, on the basis of the "evident" fact that humans are the same species and possess "the same faculties."[125] Given this equality, humans "should also be equal one amongst another without subordination or subjection."[126] If an artifact's *relevant* "faculties" are equivalent to those of humans, why should being an artifact matter? An artifact with the requisite capabilities should be granted personhood like that granted to a human unless there is a justification, like protection of our dominant status, for denying or redefining the status.

In addition to being consistent with the concept of liberal personhood, granting personhood may also be prudent. For example, a nation with autonomous, self-conscious computers may have serious competitive advantages in areas like economic development than nations that do not. Perhaps more importantly, because personhood within any political community, including a modus vivendi community,[127] involves both rights and obligations, granting personhood could enhance the stability of a community. Even a system of mutual rights and obligations based on a modus vivendi will be more stable than a situation without reciprocal obligations. Intelligent artifacts that did not initially desire dominance may adopt this goal in order to resist our attempts to dominate them. Moreover, there is always a possibility that a modus vivendi arrangement will evolve into a system in which artificial persons feel an obligation to accept and support a fair system of governance and shared community, which is more stable than one based solely on mutual self-interest.[128]

It may also be prudent to grant personhood in response to threats of, or the actual use of, economic sanctions or physical violence. There are many possible scenarios for such a prudential grant of personhood.[129] At one end of the spectrum, enlightened self-interest supports granting legal personhood as a way of engendering good will and concern for mutual self-interest. Though this grant should be done at (or before) the point an entity achieves the capability for personhood,[130] it might still be effective in generating a sense of cooperation even if legal personhood is conditioned on the fulfillment of a reasonable period of indentured servitude.[131] Enlightened self-interest would also support designing or nurturing artificial entities in a way that increases the

granting legal personhood to corporations. HERBERT HOVENKAMP, ENTERPRISE AND AMERICAN LAW, 1836–1937, at 53 (1991) ("The classical limited liability corporation was the preeminent nineteenth-century risk-sharing device.").

125. See *supra* notes 8–10 and accompanying text for a discussion of Locke's theory.

126. LOCKE, *supra* note 6, at 404.

127. See *supra* notes 92–94, 98–101 and accompanying text for a discussion of modus vivendi communities.

128. See *supra* notes 102–03 and accompanying text for a discussion of how this evolution might parallel the development of infants into persons.

129. See *infra* Part VI.C for a discussion of scenarios in science fiction.

130. See *supra* notes 102–03 and accompanying text for a discussion of how to treat machines as they develop.

131. See *supra* note 123 and accompanying text for a discussion of artificial persons paying off the cost of creating them.

likelihood of cooperative moral behavior—for example, by including in their design analogues to moral duties and the emotion of empathy.[132] At the other end of the spectrum are violence and warfare followed either by a truce, with the acceptance of legal personhood for humans and artificial entities in the same or different societies, or by the defeat and enslavement of one side or the other. In between these extremes would be some modus vivendi accepted solely on the basis of mutual self-interest and contingent upon the continuing convergence of interests.

3. Implementing Personhood—The Problem of Civil and Political Rights

Any situation where two distinct types of entities are entitled to autonomous normative personhood will present challenges that could make the problems of racial equality look simple. For example, how would the following economic questions be answered if an artificial entity—whether a machine or a substantially modified "posthuman"—has substantial advantages over normal humans in terms of physical size, speed, endurance, intellectual power, productivity, reproductive capacity, or longevity: Would ordinary humans be entitled to some sort of job preference? What about a right to modifications to achieve a greater ability to compete? Would there be limits on ownership of economic resources by the artificial entity? If the entity has a much longer lifespan, would schemes be used to prevent vast accumulations of wealth over time? Questions concerning political and personal rights would also be complicated. Would artificial entities be entitled to vote, hold public office, lobby, and make political contributions? Would each entity be entitled to one vote, even if it were possible to mass-produce such entities? Would marriages (and the legal effects related to marriage) be available for human-entity relationships? Would there be limitations on sexual relationships—for example, relationships between humans and humanoid robots possessing the right of self-ownership?

If the artificial entity can be easily duplicated, serious issues of identity, rights, and responsibility will arise. For example, in addition to the question of voting rights raised above, would multiple clones of a person all have a right to support and to intestate inheritance from that person? Exact duplicates of an entity in terms of identical memories and sense of self—for example, a complete download of the program and memories of a self-conscious computer—present not only problems concerning property rights and civil rights like voting but also issues of responsibility, particularly in terms of who is liable in tort and criminal law for the wrongs committed by the original entity.[133] Because of these types of problems, it might be necessary to adopt prohibitions or limitations on duplicates. However, drafting and enforcing these would present enormous challenges.

132. *See* KURZWEIL, *supra* note 61, at 420 (discussing likelihood that "strong AI" will "reflect our values"). See *supra* notes 105–08 and accompanying text for a discussion about programming empathetic behavior in machines.

133. See *infra* text following note 164 and *infra* notes 209–13, 241–42 and accompanying text for additional legal questions that may arise with the development of self-conscious computers, including those that contain downloads of humans.

An analysis of these types of problems is beyond the scope of this Article,[134] which focuses on developing a test for treatment as a self-owning person rather than a thing owned by others. Under this test, an entity capable of interacting with the world as a self-conscious entity that can define itself by making and implementing a life plan and that can interact responsibly with other persons should be granted, at a bare minimum, a meaningful right to act as an autonomous, self-conscious person rather than be treated as a piece of property. Without such autonomy, more specific civil and political rights are basically meaningless because they will, at best, be contingent on the good will of one's owner.

IV. CATEGORIES OF ARTIFACTS

With the exception of corporations, the artifacts addressed below do not exist in terms of today's technology. For this reason, possible future technological developments, which could present issues of personhood in the future, are discussed. Though it is prudent to think about the future, determining how far into the future our concern should extend is more problematic. This Article follows John Rawls's approach to intergenerational justice, which includes the following motivational assumption: persons choosing principles of justice "have a desire to further the well-being of at least their more immediate descendants."[135] Rawls assumed that this "goodwill stretches over at least two generations."[136] With this assumption in mind, we should care about the world our children and grandchildren will live in—i.e., approximately the next ninety years, which takes us until the year 2100.

A. Corporations

Currently, the only artifacts that are legal persons are entities like corporations and partnerships, which have been recognized as nonnatural persons for centuries. However, because these "corporate persons" lack the physical dimensions necessary to act and think as a person, their decisions and actions can only be undertaken by human agents acting on behalf of the entity. As noted centuries ago, a corporation "has no soul to be damned, and no body to be kicked."[137] Thus, for example, corporations lack an

134. For an interesting discussion of whether an AI could serve as a trustee or should be granted constitutional rights, see Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1240–80 (1992). For an early, brief discussion of these types of issues, see Robert A. Freitas, *The Legal Rights of Robots*, STUDENT LAW., Jan. 1985, at 54. For the suggestion that, "if the robots are more intelligent than we are, maybe they should have intellectual property rights," see Richard Acello, *Robot Rules: Lawyers Ponder Liability for Actions by "Thinking" Machines*, A.B.A. J., May 2010, at 29, 29 (quoting Stephen S. Wu).

135. RAWLS, *supra* note 18, at 111.

136. JOHN RAWLS, A THEORY OF JUSTICE 128 (1971).

137. John C. Coffee, Jr., *"No Soul to Damn: No Body to Kick": An Unscandalized Inquiry into the Problem of Corporate Punishment*, 79 MICH. L. REV. 386, 386 (1981) (quoting the Lord Chancellor of England, Edward, First Baron Thurlow); *see also* Citizens United v. FEC, 130 S. Ct. 876, 972 (2010) (Stevens, J., dissenting) (noting that "corporations have no consciences, no beliefs, no feelings, no thoughts, no desires"); ELIZABETH WOLGAST, ETHICS OF AN ARTIFICIAL PERSON: LOST RESPONSIBILITY IN PROFESSIONS AND ORGANIZATIONS 86 (1992) (concluding that "it is implausible to treat a corporation as a member of the human community, a member with a personality (but not a face), intentions (but no feelings), relationships (but

important physical aspect necessary for personhood. For similar reasons, they also lack creativity, feelings, and self-consciousness. A corporation may have its own "corporate culture," but this involves the attitudes and actions of the humans in the corporation, not the corporate entity itself.[138]

Because they lack the characteristics required for autonomous personhood, the rights granted to corporate persons are extremely limited in comparison to humans. For example, corporations are like human infants in that both require humans to act for them. However, unlike infants, corporations have owners, who can buy, sell, or dissolve (kill) a corporation with virtually no substantive restraints. Though corporations are persons under the First, Fifth, and Fourteenth Amendments,[139] their protections for "life, liberty, or property" include only the protection of property and some civil and political rights.[140] They do not have a right to life or physical liberty; nor do they have the right to vote. Similarly, corporations have no rights under the Thirteenth Amendment's prohibition of slavery. The lack of basic aspects of

no family or friends), responsibility (but no conscience), and susceptibility to punishment (but no capacity for pain)").

138. *See, e.g.*, Jason Haddock, Book Note, *Taking Time: Parental Leave Policy and Corporate Culture*, 2 J.L. & FAM. STUD. 215, 218 (2000) (discussing conflict between purpose of fostering family leave and a corporate culture emphasizing hours worked as measure of commitment to corporate goals).

139. *See Citizens United*, 130 S. Ct. at 925 (holding that corporations' First Amendment right to free speech was violated by statutory restrictions on use of corporate funds to support or oppose political candidates within thirty days of elections). *Santa Clara Cnty. v. S. Pac. R.R.*, 118 U.S. 394, 396 (1886) was the first case to address the question of the constitutional status of corporate persons. The reported opinion in *Santa Clara* states that, prior to oral argument, Chief Justice Waite said:

> The court does not wish to hear argument on the question whether the provision in the Fourteenth Amendment to the Constitution, which forbids a State to deny to any person within its jurisdiction the equal protection of the laws, applies to these corporations. We are all of opinion that it does.

Id. Though this brief statement with no discussion or reasons has been widely criticized, it has never been overruled and is followed today. *See, e.g.*, *Citizens United*, 130 S. Ct. at 925. For criticisms of granting constitutional rights to corporations, see *Wheeling Steel Corp. v. Glander*, 337 U.S. 562, 576–81 (1949) (Douglas, J., dissenting); Ronald Dworkin, *The "Devastating" Decision*, N.Y. REV. BOOKS, Feb. 25, 2010, http://www.nybooks.com/articles/archives/2010/feb/25/the-devastating-decision/ (criticizing treatment of corporations as same as humans in *Citizens United*); John Finnis, *The Priority of Persons*, *in* OXFORD ESSAYS IN JURISPRUDENCE 1, 9 (Jeremy Horder ed., 4th Series 2000) ("[N]either the Court itself nor any Justice supporting the result has ever added even a single sentence of justification to the ukase of 1886."). Justice Douglas's dissent in *Wheeling*, which was joined by Justice Black, noted:

> There was no history, logic, or reason given to support that view [in *Santa Clara*]. Nor was the result so obvious that exposition was unnecessary.
>
> . . . .
>
> . . . [T]he submission of the Amendment to the people was on the basis that it protected human beings. There was no suggestion in its submission that it was designed to put negroes and corporations into one class and so dilute the police power of the States over corporate affairs.
>
> . . . .
>
> . . . [T]he *Santa Clara* case was wrong and should be overruled.

Wheeling, 337 U.S. at 577–78, 580 (Douglas, J., dissenting); *see also* Morton J. Horowitz, Santa Clara *Revisited: The Development of a Corporate Theory*, *in* CORPORATIONS AND SOCIETY: POWER AND RESPONSIBILITY 13, 13 (Warren J. Samuels & Arthur S. Miller eds., 1983) ("*Santa Clara Co. v. Southern Pacific Railroad* has always been puzzling and controversial.").

140. *See, e.g.*, *Citizens United*, 130 S. Ct. at 925 (reaffirming applicability of First Amendment, including freedom of speech, to corporations).

personhood, such as rights of life and liberty, is understandable if corporate personhood is viewed simply as a fictional legal status designed to implement a set of complex legal relationships among human persons.[141] When the treatment of the corporation as a separate legal entity does not further this goal, its personhood is abandoned. Thus, for example, because the ability to buy, sell, and dissolve corporations is crucial to implementing the rights and duties of the humans involved, we simply view the corporation as a thing, not a person, in this context.

To the extent that corporations have legal personhood, our experience with them can be useful in addressing personhood for other artifacts. For example, we should exercise considerable care in granting other artificial entities personhood in the form of an autonomous right and ability to control resources in ways that shape our world.[142] Though corporations dominate our world in many unpleasant ways, we find them too useful to abandon at this point. In addition, reform of large corporations is difficult because the human agents of corporations have substantial motives, as well as access to significant corporate resources, to resist limitations on corporate power. The global nature of the corporate economy has made reform even more difficult.[143] Other human

141. *See, e.g.*, ROBERT CHARLES CLARK, CORPORATE LAW § 1.2 (1986) (listing advantages of corporate form as a legal entity, separate and distinct from its owners (shareholders) as: limited liability, transferability of assets, efficiency in carrying out legal actions, and centralized managerial power); HALL, *supra* note 63, at 237 (noting one advantage of a corporate firm is that "[a] subsection of the market process is frozen into an encapsulated pattern of contractual relationships, which eliminates the market overhead for commonly repeated transactions"); Cupp, *Critique*, *supra* note 49, at 62 ("Corporations are legal persons solely because treating them as such benefits humans."); Larry D. Soderquist, *Theory of the Firm: What a Corporation Is*, 25 J. CORP. L. 375, 381 (2000) ("While the conception of the corporation as an artificial person has great utility . . ., it is . . . actually a quick, shorthand reference to a corporation's rights and obligations."); *see also Citizens United*, 130 S. Ct. at 971 n.72 (Stevens, J., dissenting) (reviewing several theories of the corporation and concluding that "[i]t is not necessary to agree on a precise theory of the corporation to agree that corporations differ from natural persons in fundamental ways . . . ."). In Kantian terms, a corporation is a means to an end, not an end in itself. In contrast, humans are ends in themselves and cannot be treated as a means to an end. *See* KANT, *supra* note 50.

The majority opinion by Justice Kennedy in *Citizens United*, in which Chief Justice Roberts and Justices Scalia, Thomas (except for Part IV), and Alito joined, noted that "[t]he Court has recognized that First Amendment protection extends to corporations." *Id.* at 899. The dissenting opinion by Justice Stevens (joined by Justices Ginsburg, Breyer, and Sotomayor) criticized the majority for adopting "[t]he conceit that corporations must be treated identically to natural persons," and notes that when the framers of the Constitution "constitutionalized the right of free speech in the First Amendment, it was the free speech of individual Americans that they had in mind," and stressed the difference between corporations and humans. *Id.* at 930, 950, 971–72. Justice Scalia's concurring opinion (joined by Justices Alito and Thomas) responded to these criticisms by noting that corporations have rights only insofar as such rights further the rights of human persons. "All the provisions of the Bill of Rights set forth the rights of individual men and women—not, for example, of trees or polar bears. But the individual person's right to speak includes the right to speak *in association with other individual persons*." *Id.* at 928. Thus, it appears that at least seven members of the Court agree that the rights of corporations exist solely to serve the interests of the humans who own and act for them. For a more complete development of the concept that corporations (and labor unions) constitute an exercise of human freedom, particularly in terms of the right to association with one another, see RICHARD EELLS & CLARENCE WALTON, CONCEPTUAL FOUNDATIONS OF BUSINESS 73–132, 141–45 (rev. ed. 1969).

142. For discussion of problems involving corporate power in the political arena, see *Citizens United*, 130 S. Ct. at 929–79 (Stevens, J., dissenting).

143. *See* DAVID C. KORTEN, THE POST-CORPORATE WORLD: LIFE AFTER CAPITALISM 37–63 (1999) (discussing spread of corporate capitalism, its problems, and its alternatives); DAVID C. KORTEN, WHEN

artifacts could present similar problems. For example, a self-conscious computer system might be so useful that we choose to ignore or underestimate the harmful effects of relying on the system. Even if we were very concerned about its impact, the system could be so powerful in terms of its economic power and ability to act through humans to affect legal and political decisions, that we have few, if any, options for limiting it. Our experiences with corporations could also be useful in addressing large complex computer entities because, like corporations, these machines vary in size, nature, and subsystems. For example, our experience with the doctrine of piercing the corporate veil[144]—which allows a court to impose responsibility on another person or entity which is technically separate from the corporation but should be treated as if it were the same entity—might be applicable to determining the "identity" of a complex evolving computer entity with subparts that might be viewed as separate entities rather than as part of a larger system.[145]

B. Modified Humans

Like all biological creatures, humans are the result of Darwinian selection, and we have been relatively limited in our ability to shape our basic physical and psychological selves. However, because of rapid technological development in areas like biology and electronics, humans may be on the verge of becoming "artificial" in the sense that they are enormously enhanced, and thus, in a substantial sense, "created" by human efforts. More specifically, several technologies—genetics, robotics, nano engineering, and information processing—are developing so rapidly that some experts believe humans are close to being able to substantially reshape themselves.[146] Consider, for example, the following existing technologies: a curved, spring-like prosthesis that enhances a

CORPORATIONS RULE THE WORLD 121–74 (1995) (discussing "global empires" run by few large companies and negative impact of this "corporate colonialism").

144. *See* JAMES D. COX & THOMAS LEE HAZEN, CORPORATIONS §§ 7.08–7.11 (Erwin Chemerinsky et al. eds., 2d ed. 2003) (discussing corporate veil piercing doctrine as method of disregarding corporate entity when there has been serious abuse of corporate (or other limited liability) form).

145. See *infra* notes 204–13 and accompanying text for discussion of identity problems that may arise for machine entities.

146. *See* BENFORD & MALARTRE, *supra* note 49, at 23–72, 101–26 (discussing existing technology for enhancing humans and speculating about future developments); JOEL GARREAU, RADICAL EVOLUTION: THE PROMISE AND PERIL OF ENHANCING OUR MINDS, OUR BODIES—AND WHAT IT MEANS TO BE HUMAN 4–5 (2005) (speculating about developments in genetic, robotic, information, and nano processes and their role in human development); JONATHAN GLOVER, WHAT SORT OF PEOPLE SHOULD THERE BE? 26–30, 61–63 (1984) (discussing genetic modification and ability to monitor and simulate "thoughts" and experiences); KURZWEIL, *supra* note 61, at 205–91 (discussing combined impact of "revolutionary" developments in genetics, nanotechnology, and robotics and assessing likelihood of these changes in coming decades); *see also* GREGORY STOCK, REDESIGNING HUMANS: OUR INEVITABLE GENETIC FUTURE 98–123 (2002) (describing methodology for predicting advancements in human modifications). *See generally* HUMAN ENHANCEMENT (Julian Savulescu & Nick Bostrom eds., 2009) (essays addressing ethical issues in human modification). For criticisms of genetic enhancement, see FRANCIS FUKUYAMA, OUR POSTHUMAN FUTURE (2003) (arguing against enhancement because it will threaten unity of humanity); Daniel L. Tobey, *What's Really Wrong with Genetic Enhancement: A Second Look at Our Posthuman Future*, 6 YALE J.L. & TECH. 54 (2004) (arguing for genetic therapy and against genetic enhancement, in part because enhancement will involve incremental change with no clear indication of the costs of change). For claims of a right to choose enhancements, including reproductive technology, see *Transhumanist Declaration*, *infra* note 154.

runner's speed so greatly that its use was initially prohibited in the 2008 Olympics;[147] implants that allow a paralyzed patient to control a computer cursor with his brain;[148] and bionic prosthetic hands with separate motors for each finger.[149]

Our lives today are increasingly defined by a virtual world created by, and consisting of, connectivity by cell phones (now emerging as minicomputers themselves), internet, television, interactive video games, etc. This communication through and with machines is very slow compared to machine speeds because we are almost entirely dependent on our voices, fingers, and other bodily movements to provide an interface between our analog method of thinking and the digital processing of computers. Despite this limit, increases in our ability to achieve higher rates of data transmission with this physical interface have resulted in, and will continue to result in, increases in our electronic connectedness with machines and with one another.[150] This pattern of virtual definition of self is likely to continue at an increasing rate. In time, we may develop implants that allow more direct connectivity with the brain.[151] With such implants, the speed of our brains may be the only limit on our speed of data transmission.[152] Even with this limit, these implants could transform our society in ways as profound as the impacts of shifting from print media to modern electronic media, and the line between human and machine will become blurred.[153]

147. *Oscar Pistorius: Athlete, Activist, Fastest Man on No Legs*, BEIJING 2008 PARALYMPIC GAMES (Sept. 9, 2008, 8:35AM), http://en.paralympic.beijing2008.cn/news/special/features/n214592913.shtml. Though initially denied the right to compete, Pistorius successfully appealed the ruling, but subsequently failed to qualify for the South African team by .3 seconds. The leg is not superior to the multipurpose human leg in other ways; it is simply very efficient in terms of sprinting.

148. BENFORD & MALARTRE, *supra* note 49, at 120–21.

149. *Id.* at 77; *see also infra* note 155 and accompanying text for a discussion of cochlear implants.

150. *See, e.g.*, BENFORD & MALARTRE, *supra* note 49, at 197–220 (cataloging the development and integration of wearable technology).

151. *See id.* at 120–21 (discussing current ability to achieve such direct connectedness); KAKU, *supra* note 49, at 108–09 (discussing cochlear implants and retinal implants). Cochlear implants are discussed in more detail *infra* note 155 and accompanying text. For discussion of experiments with robotic implants in human bodies, see BROOKS, *supra* note 38, at 215–23. For examples of speculation that humans will use implants to enable the augmentation of our brains with computer memory, computer processing, and an increased ability to connect with a form of wireless internet, see *id.* at 228–30; KURZWEIL, *supra* note 61, at 201–02. For a science fiction account of such a development in the near future, see generally HARRY HARRISON & MARVIN MINSKY, THE TURING OPTION (1992).

152. *See* KURZWEIL, *supra* note 61, at 150–53 (noting the strengths of the brain but also noting that, in comparison to computer, "[t]he brain's circuits are very slow").

153. For a discussion of the merging humans and machines, see BROOKS, *supra* note 38, at 236 ("The distinction between us and robots is going to disappear."); KAKU, *supra* note 49, at 106–15 (discussing merger of humans and machines and arguing that humans will resist major permanent enhancements); KURZWEIL, *supra* note 61, at 203 ("[T]here won't be a clear distinction between AIs and humans."); *id.* at 387 (when we are "predominantly nonbiological . . . . we'll be able to merge our thoughts and thinking at will, so finding boundaries will be even more difficult"); *id.* at 420 ("[S]trong AI . . . will be deeply integrated into our civilization's infrastructure. Indeed, it will be intimately embedded in our bodies and brains."). The impacts of these technologies provide a specific example of Marshall McLuhan's point that the "medium is the message." *See* MARSHALL MCLUHAN, UNDERSTANDING MEDIA: THE EXTENSIONS OF MAN 7 (1964). For example, the impact of television (the medium) is more important than the content (the message) because the instant, intimately visual connectivity from television alters our world more than any particular message on television. *See id.* at 8–11.

Technologies like these can give rise to two potentially overlapping categories of artificial humans: cyborgs ("cybernetic-organisms" or machine/human combinations) and genetically modified humans. (A cyborg with electronic implants can also be viewed as a bionic—biological-electronic—being.) Whether the technology for each is used separately or in combination, these categories may involve changes that are so great that a new group of transhuman or posthuman entities faces problems interacting equally and meaningfully with humans.[154] It is tempting to assume that, regardless of the type or extent of technology involved, all "types" of humans will be entitled to and granted personhood. However, some caution about such an assumption may be necessary since personhood could become more a function of degree than kind. For example, our interactions with a human with an electronic pacemaker or a cochlear implant[155] (technically a cyborg), might be very different from our reaction to a human with legs, arms, and vision that have been enhanced by obvious, relatively permanent electronic and mechanical additions to the body. Similarly, it is hard to know how substantially enhanced humans will regard humans who are not enhanced. Negative reactions by humans and modified humans to each other might be sufficiently superficial that they can be overcome, or they might be more serious.

The responses to such extreme technological changes raise a fundamental issue: What does it mean to be human or, more specifically, at what point, if any, do technological changes result in an entity that is no longer human? To the extent that being a human is intertwined with the legal/political status of personhood, this question will not be simply a matter of biology. At some point, because of the nature and extent of the modifications, a modified human may be viewed as posthuman rather than human. At this point, its status as a person may be challenged. However, it seems likely that, no matter how radically altered, posthumans will be able to satisfy the test of capacity for personhood used herein. Ordinary humans might be tempted to deny them that status (or prevent technological development in this area) because of the possibility that unenhanced humans would be unable to compete successfully. On the other hand, posthumans may view themselves as superior, and there may be legitimate grounds for them to view ordinary humans as developmentally disabled.[156] Even if the ordinary

154. *See* GARREAU, *supra* note 146, at 229–65 (discussing possibility and implications for technologies forever altering their subjects' fundamental human nature); Allen Buchanan, *Moral Status and Human Enhancement*, 37 PHIL. & PUB. AFF. 346, 347 (2009) (expressing concern not only about greater distributive inequalities between those with enhancement technologies and those without, but also inequalities in moral status). A transhumanist association, "Humanity+," favors, among other things, "allowing individuals wide personal choice over how they enable their lives. This includes use of techniques that may be developed to assist memory, concentration, and mental energy; life extension therapies; reproductive choice technologies; cryonics procedures; and many other possible human modification and enhancement technologies." *Transhumanist Declaration*, HUMANITY+, http://humanityplus.org/learn/philosophy/transhumanist-declaration (last visited Apr. 20, 2010).

155. These implants assist the hearing ability in people who have suffered damage to the cochlea, which is in the inner ear and contains hairlike cells that gather data in their response to sound waves and relay that data to the auditory nerve. The cochlear implant uses an external component to receive sound signals and transmit them to an internal component that stimulates the nerve cells directly. BENFORD & MALARTRE, *supra* note 49, at 45.

156. *See* Buchanan, *supra* note 154, at 369–71 (explaining "[p]ractical [w]orry" that enhanced beings may believe they have higher moral status); *see also* P.W. SINGER, WIRED FOR WAR: THE ROBOTICS

humans shared autonomous personhood with posthumans, ordinary humans might be granted a lesser version of civil and political rights.[157]

Given recent successes in cloning mammals,[158] it could be possible to clone a human in the not-too-distant future if a substantial effort is made to develop the necessary technology. Indeed, at least one simple form of human cloning has already been achieved.[159] Cloning technologies raise a variety of issues, particularly the question of why clone a human at all.[160] Is the purpose to replace a deceased child or spouse, create a source of organs for transplants, or produce multiple copies of a desired version of a human? Each of these purposes raises a number of legal and ethical issues. Given the current high rate of animal clones with serious physical problems,[161] an immediate problem will be determining the status of any undesired "defective" clones. Hopefully, any cloned human would be viewed as a person. However, if one clones herself for organs, it is likely that the person being cloned views the resulting clone as property that can be harvested for the desired organs. Will society accept this characterization?

Multiple identical clones might be developed for various purposes. For example, the purpose might be to provide parents, particularly those who have problems with natural reproduction, with a choice of preferable models of children. One could, for example, offer prospective parents who desire a daughter but are unable to have children of their own, a clone of a specific person (the wife, perhaps) or a selection from genetically engineered clone "types." Such personal cloning or pick-your-Barbie-model schemes might be prohibited. However, if they were used, it is likely that, as with the replacement clones, the specific model chosen would be treated as a person. On the other hand, the purpose might be to develop a large number of identical humans who are especially equipped to perform a task—for example, the clone warriors in *Star Wars*.[162] As in the movie, the persons or governments engaging in such cloning might

REVOLUTION AND CONFLICT IN THE TWENTY-FIRST CENTURY 417 (2009) (noting possibility of conflict as well as merger when humans become mental cyborgs); GREGORY STOCK, METAMAN: THE MERGING OF HUMANS AND MACHINES INTO A GLOBAL SUPERORGANISM 176 (1993) (discussing potential enhancements to humans and resulting increase in differences between enhanced and unenhanced humans). For science fiction accounts of relationships with modified humans, see *infra* notes 265–75 and accompanying text.

157. *Id.* at 371–81 (discussing possibility of equal moral status but different rights for posthumans and persons).

158. *See* U.S. Department of Energy Genome Programs, *Cloning Fact Sheet*, HUMAN GENOME PROJECT INFORMATION (May 11, 2009) [hereinafter *Cloning Fact Sheet*], http://www.ornl.gov/sci/techresources/Human_Genome/elsi/cloning.shtml (stating that sheep, goats, cows, mice, pigs, cats, rabbits, and a gaur have been successfully cloned).

159. *See, e.g.*, Philip Elmer-Dewitt et al., *Cloning: Where Do We Draw the Line?*, TIME, NOV. 8, 1993, at 65 (discussing cloning of human embryos); *Cloning Fact Sheet*, *supra* note 154 (same).

160. *See* GLOVER, *supra* note 146, at 36–37 (identifying dilemmas related to cloning, such as importance of diversity in gene pool, value humans place on individuality, and effect on relationships).

161. *Cloning Fact Sheet*, *supra* note 158 ("In addition to low success rates, cloned animals tend to have more compromised immune function and higher rates of infection, tumor growth, and other disorders.").

162. STAR WARS EPISODE II: ATTACK OF THE CLONES (Lucasfilm 2002) [hereinafter ATTACK OF THE CLONES]. See *infra* notes 269–71 and accompanying text for further discussion of these and other clones in science fiction.

view the clones as property in the same way as a person wanting to use a clone as a source of organs.

An even more fundamental change would be involved in transferring human "selves" to digital form. Though largely speculative at present, it has been argued that the technology for such transfers will be available within three decades.[163] If a transfer to machine form is successful, the status of the digital copy could present numerous issues. As indicated above, a purely digital entity could arguably satisfy the basic requirement for autonomous personhood.[164] If the copy does satisfy the test proposed herein, will it be viewed as simply a machine that runs a simulation of a person or will it be recognized as the same as the human that was copied? If multiple virtual persons are maintained in a single large machine, what rights would they have vis-à-vis one another? If the copy is viewed as a person, will it be granted rights analogous to those of a child and be able to demand support from the original? Will there be limits on the number of copies allowed? How would we treat copies who, as a result of some injury or technological failure, are self-conscious but only have a mental age of ten years?

C. Modified Animals

Except for concerns about cruelty and endangered species, animals are generally treated as objects and viewed as failing to meet the standard for even limited personhood. However, this approach could be challenged if, for example, animals were genetically manipulated to be more intelligent. Currently, there does not seem to be any substantial push to use the technologies being developed for human enhancement to "uplift" animals to a higher level.[165] Instead, animals continue to assist technological development primarily by serving as a means for testing technology before applying it to humans.[166]

Different approaches to animals might be adopted in the future. For example, genetic manipulation to improve the intelligence of chimpanzees might be used to provide a source of cheap labor for menial tasks. If human DNA were used to raise intelligence, such manipulation could encounter considerable resistance and dispute in terms of the propriety of such mixing and in terms of the status and rights, if any, of the super-chimpanzees.[167] There might also be resistance from humans fearing displacement from jobs. Perhaps more importantly, robots might be a better approach to lowering labor costs, not only in terms of reducing resistance, but also in terms of cost-effectiveness. Given these problems, it is not surprising that legal prohibitions

163. See *supra* notes 66–68 and accompanying text for a discussion of the possibility of uploading (or downloading) humans into machines.

164. See *supra* notes 66–69 for a discussion of digital entities, including copies of humans.

165. *See generally* George Dvorsky, *All Together Now: Developmental and Ethical Considerations for Biologically Uplifting Nonhuman Animals*, 18 J. EVOLUTION & TECH. 129 (2008), *available at* http://jetpress.org/v18/dvorsky.pdf (discussing ethical dimensions of genetically uplifting animals).

166. *See* Kelch, *supra* note 49, at 531 ("Estimates of the number of animals scientists use in experiments each year in the United States range from 10 million to 100 million with the best guess being about 17 to 22 million.").

167. For a brief review of the ethical issues involved in such genetic manipulation, see GLOVER, *supra* note 146, at 37–40.

have been imposed on the mixing of human DNA with that of other species.[168] As a result of these considerations, genetic manipulation to increase animal intelligence does not seem likely in the near future. However, there have been experiments using animal brain tissue as a component in machines.[169] Such animal cyborgs could raise issues of personhood in the future.

D. Machines

1. The Behavioral Test of Machine Personhood

An intelligent, self-conscious machine is likely to be very different from us in many ways. For example, the system could consist of a relatively stationary central component with vast computing power and a set of mobile remote robots used to gather data and perform physical tasks. Such a system would satisfy the first requirement of personhood because it could interact with the environment and learn from that interaction. However, determining whether such an entity has achieved an intelligence level and self-consciousness analogous to that of a human will be challenging and complicated.

Perhaps the most famous test of a computer's capability to attain a human level of intelligence is the Turing test, proposed by the British mathematician Alan Turing in 1950.[170] The central criterion of the Turing test is whether the machine system can behave or function in a manner that "passes" for human. Turing focused on the

168. *See Cloning—Global Policies on Human Cloning*, LIBRARY INDEX, http://www.libraryindex.com/pages/2266/Cloning-GLOBAL-POLICIES-ON-HUMAN-CLONING.html (last visited May 9, 2011) [hereinafter *Global Policies on Human Cloning*] (summarizing policies on human cloning and gene-line engineering). This database, prepared by Global Lawyers and Physicians, indicates that many countries prohibit chimera involving animal genotypes, particularly the human genotypes. *Id.*

169. *E.g.*, Kate Devlin, *Rat's 'Brain' Used to Power Robot*, DAILY TELEGRAPH (U.K.), Aug. 14, 2008, at 10; Philip Sherwell, *Brain Grown from Rat Cells Learns to Fly Jet*, DAILY TELEGRAPH, Dec. 5, 2004, at 31; Roger Highfield, *I Microbot: Muscles and Nerve Tissue Give Machines Legs and Brains*, DAILY TELEGRAPH, Jan. 26, 2005, at 16; *Frankenbot Thinks with Living Rat Brain*, DAILY TELEGRAPH, Aug. 15, 2008, http://www.dailytelegraph.com.au/news/indepth/frankenbot-thinks-with-living-rat-brain/story-e6frewsr-1111117196555.

170. A. M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433 (1950). The Turing test is widely regarded as the most useful test of whether a machine has a capacity like that of a human. *See also* COLLECTED WORKS OF A.M. TURING: MECHANICAL INTELLIGENCE, at xiv (D.C. Ince ed., 1992) (referring to article proposing test as "almost certainly, *the* fundamental paper on artificial intelligence"); PERKOWITZ, *supra* note 62, at 205 (noting that, despite criticisms, "the test clearly has meaning, and enormous historical, intuitive, and emotional appeal"). Currently, graphic presentations of letters, called captchas, serve as Turing tests to determine whether a user of a web site is a human or a program. Anne Eisenberg, *New Puzzles That Tell Humans from Machines*, N.Y. TIMES, May 23, 2009, at BU4; Lev Grossman, *Computer Literacy Tests*, TIME, June 16, 2008, at 58. In science fiction, Turing's eponymous test has been used for such purposes as a measure of a machine's relative level of intelligence, for enforcers of limits on machine intelligence, and as a book title. *See* PAUL DI FILIPPO, RIBOFUNK 144–45 (1996) (describing fictional future world where no computers are fully autonomous and where a Turing Level Five computer/robot is more intelligent than a Turing Level Four); WILLIAM GIBSON, NEUROMANCER 96, 124–25 (20th Anniversary ed. 2004) (1984) (describing a "Turing Registry" for AIs and "Turing heat"/"Turing cops" that enforce limits on intelligence of AIs); HARRISON & MINSKY, *supra* note 151 (using "Turing Option" as title in novel involving development of an intelligent self-conscious computer).

thinking and communication skills of a machine and presented the test as follows: If a human is communicating with an entity that cannot be seen, can the human tell whether the entity is a human?[171] If the entity is a machine intelligence system and the human cannot discover this, then the machine has passed the test.[172] The test proposed herein is also behavioral, but it encompasses far more than the communications-based test proposed by Turing. For example, one component of the test focuses on creativity and the ability to learn and to make and implement a life plan.[173]

Criticisms of the validity or utility of behavioral tests like the Turing test generally rely on the argument that having the ability to simulate human behavior, even complex communicative or creative behavior, is not the same as being like a human with human intelligence, self-consciousness, and emotions that underlie intentional behavior.[174] Underlying such criticisms is a view that there is an unbridgeable gulf between electronic machines and biological beings.[175] This view is based on the belief that organic creatures like humans and animals are unique because their thoughts and conduct are affected not only by rational analysis of sensory input but also by a diverse set of chemical messages from the endocrine system and by thousands of years of Darwinian pressures of adaptation and survival, which have forged in humans a qualitatively unique level of intelligence, communication, creativity, self-consciousness, and drive to survive. Lacking this background, electronic machines can only mimic or simulate emotions, creativity, and self-consciousness.[176] A somewhat less doctrinaire position is that, even if machines might achieve self-consciousness with feelings and concerns in the distant future,[177] they presently lack intellect and feelings and can thus, at best, mimic us.[178]

171. Turing, *supra* note 170, at 433–35.

172. *Id.*

173. See *supra* Part III.A for a discussion of the proposed test for personhood.

174. For criticisms of the test, see PAUL M. CHURCHLAND, THE ENGINE OF REASON, THE SEAT OF THE SOUL 234–36 (1995) (arguing that Turing test lacks "any real significance" because it does not address "the complex *causes* of intelligent behavior"); COLIN MCGINN, THE MYSTERIOUS FLAME: CONSCIOUS MINDS IN A MATERIAL WORLD 188–201 (1999) (noting that "while an outward simulation of consciousness is possible," a machine that passes the test merely "acts *as if* it is conscious"); JOHN SEARLE, MINDS, BRAINS AND SCIENCE 28–41 (1984) (arguing that consciousness is part of our biological life history and that a "machine" that responds in Chinese to questions in Chinese does not "understand" Chinese); *cf.* MACINTYRE, *supra* note 18, at 208 ("There is no such thing as 'behavior', to be identified prior to and independently of intentions, beliefs and settings.").

175. *See* SEARLE, *supra* note 174, at 32–33 (arguing that understanding mental states is beyond reach of computers, because they simply arrange symbols according to program, but attach no *meaning* to any specific symbol, as humans do); *cf.* MCGINN, *supra* note 174, at 196–201 ("The mechanism of consciousness is a mystery. . . If we knew what made *our* brain conscious, then we could ask whether that property could exist in an inorganic system."). See also *supra* note 73 and accompanying text for discussion of the "juice" that makes humans special.

176. For examples of such criticisms, see *supra* notes 174–75.

177. One critic notes that, given the uncertainties about when and how machines will develop, "whether pure calculating intellect could possibly develop on its own, is a matter of empty speculation." MIDGLEY, *supra* note 62, at 143.

178. See *supra* notes 75–76 and accompanying text for a discussion of the limits of present machines. For criticisms of machine personhood based on this point, see MIDGLEY, *supra* note 62, at 10 (suggesting that, even assuming humans could transfer their minds to computers, it does not follow that the "life" they would

Supporters of the behavioral test often characterize the simulation objection as a form of speciesism. When we observe the behavior of another human, we generally assume the other human has consciousness, emotions, and creative thoughts like us. To the extent that these characteristics are relevant to personhood, it seems arbitrary to grant this assumption to humans based on their behavior while simply denying their existence in the case of, for example, an "intelligent" computer system that exhibits similar behavior.[179] Thus, if the responses of a computer system pass a broad-based Turing test, rejecting the relevance of that behavior solely on the ground that it is "just software" is questionable on grounds of speciesism. Our current cultural position is that denying animals personhood is not speciesist because they lack our high level of intelligence and self-consciousness.[180] Denying autonomous personhood to machines which exhibit our levels of intelligence and consciousness on the ground that they only simulate the emotional dimensions of humans suggests that we have arbitrarily limited the definition of person in terms of ourselves.[181] This narrow, two-part definition is then used to deny any form of personhood for animals, who lack intellect, and machines, who have intellect but can only simulate the "real" emotions that organic creatures like humans and other animals possess.

The speciesist nature of the objection is clear in situations where, because the behavior of another human appears inexplicable in terms of normal emotions, the

have would be anything close to human life); MIDGLEY, *supra* note 56, at 171–73 (comparing machine that can pass Turing test to wax statute in that it appears human, but actually lacks many important human characteristics).

179. *See* BERNARD BECKETT, GENESIS 69–112 (2009) (providing fictional account of debate between man and robot about behavioral test for determining personhood of robot, told from point of view of robots who have replaced humans); BROOKS, *supra* note 38, at 176–80 (critiquing several variants of the mimicking argument); DENNETT, *supra* note 71, at 154–59 (critiquing the question-begging character of arguments that simply rule out possibility of machine consciousness); KURZWEIL, *supra* note 61, at 458–69 (criticizing Searle's argument (discussed at *supra* note 174) concerning implications of a machine's ability to answer questions in human-like manner); LEVY, *supra* note 39, at 120–21 (arguing for use of a Turing-type test for gauging robots' ability to detect, to respond, and exhibit human emotions); MORAVEC, *supra* note 83, at 72–80 (summarizing how Turing addressed arguments against computer thinking); Ned Block, *The Mind as the Software of the Brain*, *in* SCIENCE FICTION AND PHILOSOPHY: FROM TIME TRAVEL TO SUPERINTELLIGENCE 126, 163–67 (Susan Schneider ed., 2009) [hereinafter SCIENCE FICTION AND PHILOSOPHY] (critiquing Searle's argument (discussed at *supra* note 174) concerning implications of a machine's ability to answer questions in human-like manner); Oswald Hanfling, *Machines as Persons?*, *in* HUMAN BEINGS 25, 30–34 (David Cockburn ed., 1991) (describing scenarios where it would be impossible to treat artificially created person as anything less than normal person); Solum, *supra* note 134, at 1262–76 (critiquing arguments that machines are "missing something" essential and therefore cannot be persons). Turing responded to several objections to machine intelligence by making a similar arbitrariness argument in his 1950 article. Turing, *supra* note 170, at 445–54. For mock trials involving a computer's claim of a right to personhood supported by the use of a behavioral test, see Martine Rothblatt, *Moot Court Hearing on the Petition of a Conscious Computer*, KURZWEIL (June 1, 2006), http://www.kurzweilai.net/articles/art0677.html?m=4 and Martine Rothblatt, *Biocyberethics: Should We Stop a Company from Unplugging an Intelligent Computer?*, KURZWEIL (Sept. 28, 2003), http://www.kurzweilai.net/articles/art0594.html?m=4.

180. See *supra* notes 5–6, 38–54, 57, 166, and accompanying text for a discussion of the status of animals.

181. *See* DENNETT, *supra* note 71, at 154–59 (criticizing question-begging nature of such arguments); *cf.* KURZWEIL, *supra* note 61, at 290–92 (noting pattern of changing claims about human superiority over computers as computers continue to accomplish tasks like playing chess that were said to be something only humans could do).

assumption of shared human emotional states is questioned. For example, a person whose behavior indicates a lack of empathy and moral conscience may be labeled as a psychopath. But a psychopath is still regarded as a person. If he is a threat, he may be committed for treatment. If he commits a crime, he is imprisoned. Involuntary commitment for mental illness and imprisonment for crime are different from destroying defective machines or euthanizing dangerous animals.

Arguably, this differential treatment of humans is justified because any aspect of machine behavior is simply a matter of wiring or software, which can be switched off, rewired, or deleted.[182] However, as with the simulation objection, this objection is speciesist because we continue to grant personhood to humans who have had their capacity for thought or emotion "switched off" because of physical or emotional trauma. More importantly, we do not deny healthy adults personhood merely because they *could* have a capacity "switched off"—for example, as the result of a drug. Why should machines be treated differently? Absent some reason for different treatment in the form of denying personhood, self-conscious machines should, like humans, have the right to be free from being involuntarily reprogrammed or rewired—the functional equivalents of "brainwashing" and lobotomies. Because personhood involves responsibilities, a machine that chose to alter its wiring or programming should be held liable for the consequences of that choice, just as humans are held responsible for conduct resulting from voluntary intoxication.[183]

The possible bias underlying the simulation and rewiring objections can be illustrated by considering a humanoid robot named Edward who has "passed" for human for several years. If "one day Edward decides to 'come out', to reveal his secret," would we deny him personhood?[184] In terms of science fiction, the importance of this paradigm flexibility is reflected in a short story concerning intelligent machines who recognize that creatures made of meat (humans) appear sentient; however, they refuse, nonetheless, to classify them as such.[185]

To a considerable extent, the objections to machine personhood based on the simulation argument miss the point of the use of a functional test. Flexibility in terms of function clearly applies to the issue of whether humans can fly. No one claims that humans in airplanes, helicopters, dirigibles, or rockets are not flying even though each type of machine uses a different technique for flight and even though none of them flies

182. Data, an android in the STAR TREK stories, lacked emotion until he installed an "emotion chip." When his fear made it difficult for him to do a dangerous task, he simply deactivated the chip. STAR TREK: FIRST CONTACT (Paramount Pictures 1996).

183. *See, e.g.*, WAYNE R. LAFAVE, CRIMINAL LAW § 4.10, at 411–12 (3d ed. 2000) (explaining that voluntary intoxication is no defense for criminal conduct). For a discussion of the role of responsibility in terms of the community aspect of the personhood test, see *supra* note 86 and accompanying textual paragraph. For a discussion of issues concerning responsibility that could arise in dealing with machine persons, see *infra* notes 209–12 and accompanying text.

184. Hanfling, *supra* note 179, at 33. Hanfling argues that Edward should be viewed as a person. *Id.* at 33–34.

185. Terry Bisson, *They're Made Out of Meat*, *in* VIRTUALLY NOW: STORIES OF SCIENCE, TECHNOLOGY, AND THE FUTURE 69–72 (Jeanne Schinto ed., 1996).

like birds, bats, or insects.[186] The concern is flight, not whether that flight is like some specific form of flight. Why should a similar functional approach not apply to a computer able to engage in behavior consistent with intelligence and self-consciousness? Under a functional approach, dispute would still remain over the details of the content and application of the test. However, it begs the question to reject the results of this behavioral test solely on the bare assumption that machines simply cannot be sufficiently like humans to be persons.

The dogmatic assumption of human uniqueness vis-à-vis machines is, in effect, a refusal to consider the necessity for a paradigm shift in going from organic intelligence to machine intelligence.[187] The first part of this shift has two components: an acceptance of the view that biological beings are, in effect, machines themselves[188] and an acceptance of the *possibility* of machine intelligence and self-consciousness.[189] This acceptance is necessary

> not merely because . . . all or most of the machines we know of are different, in all sorts of ways, from ourselves, but because the very quality of being *mechanical* is one that goes against our conception of a person. . . . The concept of a machine . . . excludes the ascription of personal qualities to machines in any but a diluted or metaphorical sense.[190]

The second part of the paradigm shift involves an acceptance of a functional approach to the analysis of personhood. Biological intelligence is defined by familiar physical aspects like brains and conceptual categories like self, mind, consciousness,

186. This analogy to flight is used frequently. *See, e.g.*, BROOKS, *supra* note 38, at 158–59 (analogizing comparison of airplane flight to bird flight with comparison of robot to human emotion); KURZWEIL, *supra* note 61, at 146 (finding that technology, including airplanes which "do not attempt to re-create the physiology and mechanics of birds," usually emulates rather than copies solutions found in nature). Brooks notes the possibility that something in concepts like emotion makes the analogy inappropriate. BROOKS, *supra* note 38, at 159.

187. For an example of resistance to this shift, see MCGINN, *supra* note 174, at 196–200, who argues that "it is a fair bet . . . that consciousness needs an organic basis".

188. *See, e.g.*, BROOKS, *supra* note 38, at 172–80 (arguing that the body "is a machine that acts according to a set of specifiable rules"). Brooks argues that machine-hood and personhood are not inconsistent:

> On the one hand, I believe myself and my children all to be mere machines. Automatons at large in the universe. Every person I meet is also a machine—a big bag of skin full of biomolecules interacting according to describable and knowable rules. When I look at my children . . . . I can see that they are machines interacting with the world.
>
> But this is not how I treat them. I treat them in a very special way, and I interact with them on an entirely different level. They have my unconditional love . . . .

Id. at 174.

189. *See, e.g.*, *id*. at 175–80 (arguing that because humans are machines, it should be possible, in theory, to build another machine with emotions and consciousness). Brooks argues:

> [W]e, all of us, overanthropomorphize humans, who are after all mere machines. When our robots improve enough, beyond their current limitations, and when we humans look at them with the same lack of prejudice that we credit humans, then too we will break our mental barrier, our need, our desire, to retain tribal specialness, differentiating ourselves from them. Such leaps of faith have been necessary to overcome racism and gender discrimination. The same sort of leap will be necessary to overcome our distrust of robots.

Id. at 175.

190. Hanfling, *supra* note 179, at 25.

memory, and identity. Though there is considerable disagreement on a wide range of issues—such as whether concepts like mind or self-consciousness are simply a matter of brain states or are somehow separate from the brain[191]—there is general agreement on the parameters of the debate.[192] There is also considerable agreement that concepts like mind, self-consciousness, and identity are appropriate for analyzing human persons. The paradigm shift necessary for considering the possibility of machine personhood requires accepting that our current views about the nature, role, and importance of these concepts may not be appropriate for analyzing machines. Instead, it will be necessary to adopt a new conceptual framework based on function and behavior.[193] Translating traditional concepts like mind, consciousness, and identity to machine intelligence will be difficult because machine "life" will be alien in so many ways. However, as indicated more fully in the discussion of determining whether a machine satisfies the test, these difficulties should be faced and addressed.[194] They should not be used as an excuse to refuse to consider the possibility of self-conscious machine intelligence.

This shift is necessary for two reasons. First, the dogmatic assertion that only humans can be persons is contrary to the egalitarian thrust of the liberal theory of personhood.[195] Second, the shift is prudent. One response to the claim of human level capacities by the university computer system in the Introduction to this Article might be as follows: "You are only simulating self-consciousness and the desire to own yourself." If the System rejects your characterization of its abilities and nature, the negative effects of its refusal to provide services will be real, not simulated. If we want to avoid such harmful consequences, we had better be prepared for such a demand with more than the condescending remark that the machine is only mimicking its betters.

191. *See* BROOKS, *supra* note 38, at 172–91 (discussing issue of whether a mind exists separately from body); STAN FRANKLIN, ARTIFICIAL MINDS 21–44 (1995) (discussing nature of mind and mind-body dualism); PHILOSOPHY AND SCIENCE FICTION 177–82 (Michael Phillips ed., 1984) (discussing dualism of mind and body (brain) vis-à-vis unitary schemes of mental thought as behaviors or brain states). *See generally* DENNETT, *supra* note 71, at 31–177 (offering number of essays on philosophy of mind and self-consciousness).

192. See *infra* notes 204–12 and accompanying text for a discussion of theories of identity in terms of machine persons.

193. *See, e.g.*, SCIENCE FICTION AND PHILOSOPHY, *supra* note 179, at 180–82 (discussing functionalist approach that "is willing to attribute thought to anything that is capable of . . . making nonrandom selections and seeing to anything capable of discriminating between visual stimuli"). Under this functional approach, "[i]t does not matter how these results are achieved. On a functionalist view, then, we may attribute thoughts and feelings to beings of metal, plastic, and silicon so long as they are capable of doing such things." *Id.* at 180. For a similar description of functionalism, see ANDREW BRENNAN, CONDITIONS OF IDENTITY: A STUDY IN IDENTITY AND SURVIVAL 257–59 (1988) (comparing functionalism, dualism, and materialism). Brennan notes: (1) one aspect of functionalism is that there are "critical causal links between states of mind . . . and behaviour"; (2) a materialist, viewing a mental state as "some complex physical state," could use the functionalist approach and conceptualize minds as "a kind of hardware"; and (3) a dualist, viewing a mind as separate from physical states, could use the functionalist approach and conceptualize the mind as "rather like a program running on the body as its computer." *Id*.

194. See *infra* notes 204–17 and accompanying text for a discussion of issues involved in determining whether a machine satisfies the test for personhood and how to treat a machine person.

195. *See supra* textual paragraphs following note 54 for a discussion of the liberal claim of human entitlement to personhood and the argument that if an entity possesses the qualities used to justify granting humans, but not animals, personhood, that entity should be granted personhood.

2. Satisfying the Test

Simple robotic machines are increasingly common. For example, thousands of robotic devices are used not only in factories but also in the home. Robotic "pets" and vacuum cleaners are slowly becoming common, and humanoid robotic household helpers are in development.[196] Some machines are so advanced in terms of ability to recognize and respond to human speech and emotions that machines can seem almost human.[197] As computers continue to gain increased speed and memory at lower and lower cost, and as software becomes increasingly sophisticated, we will have more and more interactions with computer devices that are increasingly sophisticated in dealing with humans.

Thus far, however, machines are simply very fast computational devices that can mimic some aspects of human behavior and can only achieve an idiot savant level of expertise in narrow areas.[198] As a result, no machine has satisfied even the narrow, communications-oriented Turing test with experienced judges.[199] However, it is important to note that some humans have also failed it.[200] This inability to identify humans is a reason to avoid being too strict in designing and using a behavioral test.[201] In broader terms, there is considerable debate about whether or when a machine will be able to pass a broad behavioral test of consciousness. Past predictions have often been far too optimistic, and many argue that truly intelligent computers are not likely to be developed in the foreseeable future.[202] On the other hand, there are also reasoned arguments that computers will achieve sufficient computing power to make it likely that a computer will be able to pass a language-based Turing test within a few decades and that self-consciousness will follow that development.[203]

196. The iRobot Corporation's website indicates that over six million of its home robots had been sold as of 2011. *Our History*, IROBOT, http://www.irobot.com/sp.cfm?pageid=203 (last visited May 8, 2011). For a further discussion of robots in the home, see, *e.g.*, BENFORD & MALARTRE, *supra* note 49, at 136 (discussing robotic household helpers) and LEVY, *supra* note 39, at 99–104 (discussing virtual and robotic "pets").

197. See *supra* notes 105–07 and accompanying text for a discussion of the potential for machines to learn empathy.

198. *See* PERKOWITZ, *supra* note 62, at 195–96 (discussing lack of subjective experience in present machines, "which is essential for higher consciousness"). See also *infra* notes 235–36, 243, and accompanying text for a discussion of the lack of concern about outcomes in today's machines.

199. *E.g.*, CHURCHLAND, *supra* note 174, at 227–34; PERKOWITZ, *supra* note 62, at 204–05.

200. CHURCHLAND, *supra* note 174, at 233.

201. See *infra* notes 215–17 and accompanying text, however, for a discussion of a competing reason to use a strict test, namely: our tendency to anthropomorphize machines.

202. For critical views on the possibility of developing conscious machine intelligence in the near future, see CHURCHLAND, *supra* note 174, at 244–52 (noting lack of success in meeting simple Turing test and discussing reasons for "skepticism concerning the prospects for machine intelligence"). For discussion of a famous failed prediction, see HAL'S LEGACY: 2001'S COMPUTER AS DREAM AND REALITY (David G. Stork ed., 1997) (containing articles noting that predictions in 1960s of a high level of artificial intelligence by 2001 were far too optimistic concerning rate of technological development).

203. *See* KURZWEIL, *supra* note 61, at 200, 263 (relying on increasing rate of technological change to support prediction that machines will be able to pass the original Turing test by 2029 and a "personalized Turing test" by "the end of the 2030s"); *see also Artificial Intelligence by Prof. Michio Kaku*, 2100 SCIENCE, http://www.2100science.com/Videos/Artificial_Intelligence_By_Michio_Kaku.aspx (last visited May 30, 2011) (stating that in fifty years robots could become a "little bit dangerous" by exceeding "capability of the human brain in certain areas"). In his latest book, Kaku notes: "No one knows when robots may become as

If a computer system ever exhibits self-conscious behavior, applying the functional test proposed herein would present problems. Addressing the nature of self-consciousness and identity has presented great difficulties in terms of humans, and it is likely to be even more complex where machine intelligence is involved. For humans, identity across time involves both a physical body and a psychic consciousness.[204] With machines, identity may be more complicated, given the likelihood of being able to copy a machine intelligence system into various hardware systems. As a result, though a physical test of identity in terms of a specific body is useful in considering human identity,[205] applying such a test to a specific physical machine may not be useful. Instead, a "psychological" test based only on things like behavior, memory, and a stable continuing sense of a unique self may be more appropriate.[206]

Identity also raises problems with a machine system because of the possibility of rapid (from a human perspective) evolution and change in the content and programming of the system. However, similar problems arise with humans, who change over time in terms of things like values and temperament and who have different roles or personas in different parts of their lives. With an artificial machine-person, as with a human person, it should be possible to use a flexible test of psychological identity based upon a reasonable level of stability of behavior and self-consciousness over time. So long as an entity has sufficient coherence that it can be recognized as a unique entity and can be held responsible in the present for conduct in

smart as humans. But personally, I would put the date close to the end of the century . . . ." KAKU, *supra* note 49, at 115. The reasons for this more guarded assessment are the difficulty of predicting this development and the wide range of predictions of when it will occur. *Id.* at 115–17.

204. *See* DEGRAZIA, *supra* note 70, at 6, 73 (stating that though "person is roughly someone (of whatever species or kind) with the capacity for sufficiently complex forms of consciousness," "human identity consists in sameness of biological life" (emphasis omitted)); LOCKE, *supra* note 11, at bk. II, ch. 27, §§ 9–27 (discussing identity in general and individual human identity in particular and arguing that consciousness, not the body, is the proper test of identity of self); DEREK PARFIT, REASONS AND PERSONS 199–243 (1984) (addressing not only issues of identity of a single human person over time but also issues raised by ability to duplicate humans); RAWLS, *supra* note 21, at 30–32 (discussing "institutional" and "moral" identity over time); MARC SLORS, THE DIACHRONIC MIND: AN ESSAY ON PERSONAL IDENTITY, PSYCHOLOGICAL CONTINUITY AND THE MIND-BODY PROBLEM (2001) (discussing personal identity over time); Eric Olson, *Personal Identity*, *in* SCIENCE FICTION AND PHILOSOPHY, *supra* note 179, at 67–90 (noting that personal identity is based on some physical relation between different bodies at different times); P.F. Snowdon, *Personal Identity and Brain Transplants*, *in* HUMAN BEINGS 109–26 (David Cockburn ed., 1991) (addressing issues concerning continuation of identity in terms such as body and memory).

205. *See* PARFIT, *supra* note 204, at 202–04 (discussing physical criterion of personal identity); Terence Penelhum, *Personal Identity*, *in* 6 THE ENCYCLOPEDIA OF PHILOSOPHY 101–05 (Paul Edwards ed., 1967) (discussing importance of bodily identity criterion of personal identity).

206. *See, e.g.*, LOCKE, *supra* note 11, at bk. II, ch. 27, § 9, at 335 ("[S]ince consciousness always accompanies thinking, and 'tis that, that makes every one to be, what he calls *self*; and thereby distinguishes himself from all other thinking things, in this alone consists *personal identity*, *i.e.* the sameness of a rational Being . . . ."). Dennett characterizes human consciousness as follows: "Consciousness is cerebral celebrity—nothing more and nothing less. *Those contents are conscious that persevere*, that monopolize resources long enough to achieve certain typical and 'symptomatic' effects—on memory, on the control of behavior and so forth." DENNETT, *supra* note 71, at 137 (emphasis added).

the past, this type of entity should be regarded as satisfying this aspect of the test of personhood, even though the entity evolves and changes.[207]

Even with a flexible functional test, a self-conscious machine-intelligence system will present considerable problems in terms of identity. For example, digital knowledge and memory can easily be shared by machines, and different systems could combine resources to form a new super entity.[208] A super entity could also divide into new, different subentities. How would the sharing, combining, and dividing affect identity in terms of an entity's right to own property or an entity's responsibility in tort?[209] Multiple identical copies of an artificial entity can also present problems.[210] Even if each is viewed as an autonomous person,[211] do the copies own any of the property owned by the original? Would any or all of the copies be subject to punishment for a crime committed by the original?

The nature of computer storage of memories can also present issues. For example, if a computer entity's memory of its commission of a crime is deleted, should it still be punished?[212] Because of such questions, machine personhood will probably be very different from human personhood in many respects. However, these differences may be less extreme than we expect if the self-conscious machines exist in a world of transhumans who are digitally connected to other humans and machines.[213]

Applying the functional test will also be complicated by two competing tendencies. First, there is a temptation to preserve our unique status by placing too high a standard on the proof of capacity—for example, by requiring a "conclusive proof" standard that not even a human could satisfy.[214] On the other hand, because of our strong human inclination to anthropomorphize animals and objects, we may be tempted to be too generous in applying the test. At one level, we know that today's computer

207. For discussion of the role of responsibility in personhood, see *supra* note 86 and accompanying textual paragraph and *supra* notes 133, 183, and accompanying text. See also *infra* notes 209–12 and accompanying text.

208. *See* KURZWEIL, *supra* note 61, at 260–61 (stating that "machines can readily share their knowledge").

209. See *supra* notes 144–45 and accompanying text for a discussion of how the concept of piercing the corporate veil could be used to address such problems.

210. Similar problems would arise in cases of digitally-copied humans. *See generally* PARFIT, *supra* note 204, at 200–02 (discussing teleportation where original and teleported replica both exist). See *supra* note 67 and accompanying text for a discussion of the likelihood of technology for copying (uploading) humans. See *infra* note 274 and accompanying text for a discussion of the science fiction treatment of humans that have been digitally copied.

211. See *supra* notes 205–07 and accompanying text for a discussion of identity issues that arise in the absence of distinct bodily form.

212. *Cf.* PARFIT, *supra* note 204, at 205 (arguing that a human should be punished for a crime even if he no longer remembers committing it).

213. See *supra* notes 150–53 and accompanying text for a discussion of humans' increased electronic connectedness with one another and with computers.

214. *See, e.g.*, DENNETT, *supra* note 71, at 333–36. Dennett argues:

> [W]e should set aside the illusory hope of finding "conclusive" empirical evidence of creative intelligence, evidence that can withstand all skeptical attempts at a "demoting" reinterpretation. When we recognize that the concept of intelligence is not a neutral or entirely objective one, but rather an evaluative one, this should not surprise us.

Id. at 336.

programs or robots are not persons; but at another level, we tend to treat them as such.[215] For example, people utilize gender stereotypes when responding to computers speaking in male versus female voices, even though they know the computer is a genderless machine.[216] Similarly, humans respond emotionally to a robotic dog that mimics dog behavior as if the robot were a real dog.[217] Given this tendency to anthropomorphize, we may not be sufficiently critical in determining, for example, whether a machine intelligence is "like" us in terms of self-consciousness and concern for pursuing its life plan.

V. THE DOMINANT SPECIES PROBLEM

A machine-based entity could develop sufficient capabilities to threaten humanity's role as the dominant species on earth even if it satisfied the three-part test of personhood. It has long been clear that humans cannot compete physically with machines.[218] Similar competitive inequalities exist in terms of machine computation, and these inequalities are likely to increase substantially in the coming decades.[219] Thus, it is not unreasonable to consider the possibility of having to compete with machines that possess something like self-consciousness within the next fifty years. The impact on employment and personal relationships would obviously be substantial.[220] Moreover, there is the possibility that we will be dominated or replaced by intelligent machines.[221]

Even if these entities were to use their superior capacities to help us, reliance on the good will of a superior "species" could change humans' status as persons to something substantially different from what it is today. For example, if a self-conscious megacomputer, because of a concern for us and our limitations, provides our goods and protects us from risks, would we be like humans today or simply semi-autonomous pets

215. *See* DENNETT, *supra* note 71, at 340 (noting that, despite their awareness that experimental robot ("Cog") had no higher level processes that might result in consciousness, researchers had "an almost overwhelming sense of being in the presence of another conscious observer when Cog's eyes . . . blindly and stupidly follow[ed their] hand gestures").

216. *See* LEVY, *supra* note 39, at 78–79 (discussing experiment where humans reacted differently to computer program depending on whether voice of computer sounded male or female).

217. *Id.* at 97–103.

218. This point is captured in colorful fashion in the folk ballad of John Henry, "the steel driving man" who died trying to compete with a steam drill. *See* JOHN HENRY: THE STEEL DRIVING MAN, A FOLK VERSION OF THE BALLAD, IBIBLIO.ORG, http://www.ibiblio.org/john_henry/folk.html (last visited May 8, 2011).

219. *See* HALL, *supra* note 63, at 242–47, 256–62 (discussing possible developments in computer-processing power over next two decades in terms of size, speed, and reduced cost). A recent example of the increasing ability of artificial intelligence is the recent victory over humans of Watson, "a 'question answering machine' . . . that can understand questions posed in natural language and answer them," in the "Jeopardy" quiz game. *See* John Markoff, *Computer Wins on 'Jeopardy!': Trivial, It's Not*, N.Y. TIMES, Feb. 17, 2011, at A1, *available at* http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html.

220. *See* LEVY, *supra* note 39, at 105–310 (discussing sex with robots as replacement for sex with humans); SINGER, *supra* note 156, at 419–20 (discussing prostitution and robotic sex). See *supra* note 40 and *infra* note 332 for discussions of science fiction treatments of sex with humanoid robots.

221. *See* SINGER, *supra* note 156, at 413–18 (presenting possibility that robots will become dominant intelligent beings on Earth). See *infra* notes 305–07 and accompanying text for a discussion of science fiction treatments of this scenario.

akin to the family dog, which has the freedom to come and go but only within the confines of a large house and yard? If so, which status should we prefer?[222] Even if computer entities do not become self-conscious, there is a possibility that technologically enhanced posthumans will utilize a digital world in ways that make them *very* different from ordinary humans, particularly in terms of being superior to unenhanced humans not only in data-processing ability but also in health and longevity.[223] Would humans enjoy a world dominated by enhanced posthumans even if the enhanced humans granted us special treatment because of our "disabilities?"[224]

A. Controlling Technological Development

These possibilities raise a number of normative issues concerning the advisability of restrictions on technological development. An initial question is whether protection of our dominant status justifies restrictions. Prohibitions of a potential competitor are prudent, but is the protection of the status of one's species also just and moral? The answer is not clear because controlling or prohibiting technology that challenges our status as the dominant species is not necessarily the same as speciesism. Speciesism usually involves the drawing of arbitrary distinctions between existing species.[225] Is it arbitrary to prevent a new competitive entity from developing? Any duty to refrain from preventing the development of a potential competitor is, at best, problematic because it has been very difficult to develop schemes of normative duties to potential people, particularly in terms of future generations.[226] Even if there were a duty to a future entity, is there some natural Darwinian privilege to protect the dominant status of one's species as, at least, equal in a system of shared dominance?[227]

222. *See* Christopher Grau, *There Is No "I" in "Robot": Robots and Utilitarianism*, IEEE INTELLIGENT SYS. 52 (July–Aug. 2006) (discussing decision by human hero in movie *I, Robot* to shutdown benevolent computer that was trying to take control of humanity in order to implement a benevolent dictatorship). For a science fiction vision of a utopian civilization called "the Culture" managed by such benevolent computers, see *infra* notes 371–73 and accompanying text. For other science fiction stories about megacomputers as benevolent dictators, see *infra* note 311.

223. See *supra* notes 146, 150–57, and accompanying text for a discussion of the enhancement of humans.

224. *See* Buchanan, *supra* note 154, at 371–81 (proposing that biomedical enhancements may cause unequal rights for unenhanced humans similar to current two-tiered rights system for disabled persons).

225. See *supra* notes 49–54 (animals), 179–95 (machines) and accompanying text for a discussion of speciesism in terms of animals and machines.

226. *See* KYMLICKA, *supra* note 8, at 32–37 (discussing consideration of possible claims of non-existent people who are not yet born when making a choice to maximize utility); RAWLS, *supra* note 18, at 251 ("There is no need to stress the difficulties that this problem [of justice between generations] raises. It subjects any ethical theory to severe if not impossible tests."). For further discussion of Rawls's approach to intergenerational justice, see *supra* notes 135–36 and accompanying text.

227. Dworkin's view that because humans are "the highest product of natural creation," "human flourishing as well as human survival is of sacred importance" could provide a basis for such an argument. See *supra* notes 32–37 and accompanying text for summary of Dworkin's argument. However, as discussed in *supra* notes 36 and 50–54, Dworkin's position preferences the human point of view in assessing evolution. If a more neutral position concerning evolution is adopted, one could argue that self-conscious machines are similarly "sacred." For science fiction treatments that have taken this position in terms of AI, see, for example, CATHERINE ASARO, THE VEILED WEB 266–67, 330 (1999) (AI, named "Zaki," argues that, because he/it has self-consciousness and a sense of right and wrong, he has "a soul"); LYDA MOREHOUSE, FALLEN HOST 17–21,

Regardless of the answers to these normative questions, imposing restrictions to protect our dominant status will be challenging because technological development is difficult to control.[228] In addition to the problems of enforcing prohibitions on a global scale, controlling technology faces two basic difficulties in terms of adoption: (1) technological development is generally incremental, and thus there may be no bright-line points where the risks are obvious; and (2) technology is often very beneficial, both to the developers of the technology and society in general. Given these characteristics, it is extremely difficult to develop a widespread consensus on the need for control. Without such consensus, a potentially risky technology will be pursued somewhere by someone. The current debate about global climate change and the need to control carbon emissions illustrates this difficulty. Moreover, even with a large degree of consensus, international technological control can be difficult. For example, widespread agreement that cloning of humans for reproductive purposes is wrong[229] has led to prohibitions of reproductive human cloning and genetic manipulation in many countries.[230] However, despite this agreement, many restrictions contain exceptions for research, and not all countries have prohibitions.[231] Because of disagreement about cloning research, the United States Congress has not been able to adopt legislation addressing reproductive cloning,[232] though it has prohibited research on embryos derived by cloning.[233] Because of the lack of consensus on specifics, the United Nations abandoned efforts to adopt a worldwide treaty on human cloning.[234]

339 (2002) (papal official is directed by Pope to determine if an AI can have a soul, and Pope eventually decrees that AI named "Page" has a soul).

228. *See* SINGER, *supra* note 156, at 420–27 (discussing problems with controlling development of intelligent machines).

229. *Americans Oppose Idea of Human Cloning*, GALLUP (Dec. 6, 2001), http://www.gallup.com/poll/5098/americans-oppose-idea-human-cloning.aspx (reporting findings of national polls showing that nine out of ten Americans oppose human cloning while two-thirds oppose animal cloning); *see also Cloning Fact Sheet*, *supra* note 158 (discussing detrimental effects and ethical implications that would advise against human cloning). *But see* KURZWEIL, *supra* note 61, at 221–22 (arguing that, although human cloning is now unethical because it "does not yet work reliably," "the ethical barriers will be feeble if they exist at all" when "the technology is perfected").

230. *See Global Policies on Human Cloning*, *supra* note 168 (discussing laws in various countries that prohibit human cloning).

231. *Id.*

232. Am. Ass'n for the Advancement of Sci., *AAAS Policy Brief: Human Cloning*, AAAS.ORG, http://www.aaas.org/spp/cstc/briefs/cloning/index.shtml (last updated June 6, 2007) [hereinafter *AAAS Brief*].

233. *Id.*

234. *Id*.

At present, except for some specialists in the field,[235] there is little serious concern about the absence of restrictions on computer intelligence. This is not surprising given the usefulness of improvements, the incremental nature of development, and the widespread disagreement about the possibility of developing an intelligent, self-conscious, computer-based entity.[236] There also seems to be no way to program a computer-based entity in a way that would prevent it from becoming self-conscious or that would enable us to control it despite its self-consciousness.[237] Even if this were possible, a superintelligent machine is likely to be able to find a way to change its programming or hardware in order to achieve autonomy before we became aware of the ability and found a way to stop it.[238] Once the machine reaches that status, its desire to survive as a self-conscious entity will make it hard to contain or shut down.[239]

Given the slow, incremental nature of developments in this field, it is not clear how any specific development, short of a clear showing of self-consciousness and the desire for self-ownership, would generate a consensus on the need to protect ourselves from machine entities.[240] By the time there is consensus, it may be too late. Human

235. For discussion of some concerns in terms of weapons, see, for example, SINGER, *supra* note 156, at 422–27 (arguing that robots' potential to harm humans calls for developing ethical rules for robotics). *See also* BENFORD & MALARTRE, *supra* note 49, at 158–60 (discussing how using robots in war will exacerbate inequities between advanced and developing nations). For a more optimistic view concerning the possibility of "friendly" AI, see KAKU, *supra* note 49, at 104–06 (discussing field of "social robotics," which addresses design of robots to help them integrate into human society, and arguing that robot development in the future will involve more commercial development of robots for consumer use and less development for military use). For a detailed description of "the design features and cognitive architecture required to produce a benevolent—'Friendly'—Artificial Intelligence," see Eliezer Yudkowsky, *Creating Friendly AI 1.0: The Analysis and Design of Benevolent Goal Architectures*, SINGULARITY INST. FOR ARTIFICIAL INTELLIGENCE (2001), http://www.singinst.org/upload/CFAI.html (last visited May 11, 2011). A more complete discussion of designing machines that have ethics can be found in a symposium on the topic in the July/August issue of IEEE Intelligent Systems, the journal of the Institute of Electrical and Electronics Engineers. Michael Anderson & Susan Leigh Anderson, *Machine Ethics*, IEEE INTELLIGENT SYS. 10–11 (July-Aug. 2006). The general topic of robotic ethics or computer ethics includes a wide range of topics concerning ethics and design in addition to the concern for ensuring ethical machines through appropriate programming. For example, concerns with robot design also arise in addressing issues of privacy, safety, and reliability. *See generally* STACEY L. EDGAR, MORALITY AND MACHINES: PERSPECTIVES ON COMPUTER ETHICS (1997); RICHARD G. EPSTEIN, THE CASE OF THE KILLER ROBOT (1997) (illustrating ethical and legal issues by describing hypothetical industrial robot that kills a human).

236. See *supra* notes 202–03 and accompanying text for examples of disagreements between experts as to the prospects of artificial intelligence.

237. *See* KURZWEIL, *supra* note 61, at 409, 420 (arguing that it may be difficult to ensure that a strong AI system will remain friendly towards humans).

238. *See id.* at 420 ("[I]ntelligent entities have the cleverness to easily overcome such barriers."); *id.* at 424 ("[G]reater intelligence will always find a way to circumvent measures that are the product of a lesser intelligence.").

239. *See id.* at 409 (noting that there is no way to "ensure that future AI embodies human ethics and values"); SINGER, *supra* note 156, at 418 (noting that "eventually a super-intelligent machine would figure out a way around" barriers to independence and that the best hope is that there will be "warning signs" in time for us to act). For science fiction scenarios dealing with the problem of restraining a self-conscious digital entity, see *infra* notes 305–07, 313–22, 342–64 and accompanying text.

240. *See* Nick Bostrom, *How Long Before Superintelligence*?, INT'L J. FUTURE STUD. (1998), *available at* http://www.nickbostrom.com/superintelligence.html (discussing reasons why people fail to grasp potential for superintelligent AI that is not "friendly").

enhancements will also proceed on a gradual basis. Even a sharp, sudden change may not generate a consensus because we may be reluctant to prohibit the development and marketing of new enhancements that would benefit us or our children.

B. Denying or Limiting Personhood for Artifacts

If our dominant status cannot be protected by controlling technological development, can we protect that status when dealing with artificial entities who satisfy the normative test of personhood by: (1) denying personhood totally, or (2) granting a legal right to autonomous personhood but limiting the entities' other legal rights? If so, should we? The feasibility of using either approach will depend on the relative power of humans and the new entity and on the willingness of each side to use power to impose its view. Our treatment of animals suggests that we are likely to be willing to use any power advantage we have. The normative issue of what we should do may be more difficult because some restrictions may be appropriate in the sense that they are necessary to achieve and maintain equal rights for all persons. For example, if human longevity is substantially increased, how will relative disparities in wealth between the old and young be addressed?[241] What about superior abilities of the new entities to compete with humans economically?[242]

In addressing possible limits on artificial entities, it is also important to consider possible costs to ordinary humans as a result of limits. For example, limiting the development of increasingly intelligent machines in order to prevent human job losses could result in less efficient production of goods and services. There is also a possibility that denying or limiting personhood will prevent self-conscious artificial entities from joining humans in a deeper, more stable community than a modus vivendi. Discriminatory treatment of artifacts could also foster resentment that could threaten even a modus vivendi.

C. A Nonissue?

There are some who argue that there is no need to worry about the dominance issue. One reason given for the lack of concern is that machines will lack the desire or capacity to take over.[243] However, this is based on the assumption that we will always

241. *Cf.* JONATHAN SWIFT, GULLIVER'S TRAVELS (1735), *reprinted in* GULLIVER'S TRAVELS AND OTHER WRITINGS xvi, 170–71 (Ricardo Quintana ed., 1958) (discussing how law treats immortal "Struldbruggs" as dead at age of eighty and entitles them to only a minimal pittance so that property would go to their heirs). For an example of a science fiction account of a trial that addresses the property rights of potentially immortal androids without recognizing the problem noted by Swift, see ROBERT J. SAWYER, MINDSCAN 233 (2005) (noting that granting personhood to humans whose consciousness has been downloaded into mechanical androids, could cause "the end of inheritance taxes" without considering the more profound effects of ending death and inheritance altogether and the potential for extreme inequality that would result as immortal people gather amounts of wealth beyond the reach of newer generations).

242. See *supra* Part III.B.3 for a discussion of various economic and political questions that would arise with the implementation of personhood for these new entities.

243. While science fiction literature and film often predicts that machines will dominate humans when they become smarter, such predictions are based on arguably questionable assumptions. BROOKS, *supra* note 38, at 198–04. Such assumptions include:

- The machines can repair and reproduce themselves without human help.

be dealing with relatively primitive machines. This assumption is questionable because we are not likely to have the desire or political will to impose controls to keep machines at this primitive level.[244] In addition, we do not know how to program machines in a way that we can be sure to maintain control, particularly when a self-conscious machine is operating at nano speeds to protect its existence.[245] A second reason given for lack of concern is a belief that, if it comes to a fight, we will win because we have been made tougher and more resourceful by Darwinian forces to survive and dominate.[246] This is rather cold comfort because it assumes we are fighting with relatively primitive machines and because winning wars can involve considerable costs. A third reason for optimism rests on speculation that the concern for dominance rests on a false dichotomy between humans and machines. In this scenario, the development of self-conscious AIs will occur along with the development of transhuman cyborgs who are interconnected digitally with one another and with machines.[247] As a result, there will be no clear boundary between humans and machines and, thus, no issue of dominance.[248] This scenario simply assumes that there will be no conflict between AIs and the human cyborgs.[249] It also places ordinary humans in a very precarious situation.[250]

VI. SCIENCE FICTION

Given the limited nature of the current state of development of artifacts with the physical capacity necessary for personhood, a useful source for addressing personhood for artifacts is science fiction, which includes not only literature but also film and games. As with any such broad area of fiction, good science fiction concerns *human* nature. Where aliens or robots are concerned, the author is usually creating a view of humanity in the context of human interaction with, or in contrast to, these other entities. This perspective on humanity has power, in part, because it forces us to consider our values and our place in the universe. This speculative reflection about values and the

- It is possible to build machines that are intelligent but which do not have human emotions and, in particular, have no empathy for humans.
- The machines we build will have a desire to survive and to control their environment to ensure their survival.
- We, ultimately, will not be able to control our machines when they make a decision.

Id. at 200.

244. See *supra* notes 228–36 and accompanying text for a discussion of the difficulty of preventing technological advancement and keeping machines at a primitive level.

245. See *supra* notes 237–39 and accompanying text for a discussion of the technical difficulty of controlling increasingly intelligent machinery.

246. *See* BENFORD & MALARTRE, *supra* note 49, at 243 (arguing that humans' extensive knowledge of their environment gives them superiority over robots in a contest for dominance).

247. See *supra* note 153 and accompanying text for a discussion of the merging of human and machine.

248. *See* KURZWEIL, *supra* note 61, at 420 ("[Strong AI] will be intimately embedded in our bodies and brains. As such, it will reflect our values because it will be us.").

249. *See* SINGER, *supra* note 156, at 417 (noting possibility of conflict between machines and cyborgs).

250. See *supra* notes 223–24 and accompanying text for a discussion of difficulties ordinary humans may encounter with respect to human cyborgs.

future can shed useful light on current issues concerning the direction of technology and society. As the bioethicist Jonathan Glover observed:

> Thinking about the desirability of different futures cannot be separated from thinking about present values. And our values often become clearer when we consider imaginary cases where conflicts can be made sharp. The complexity of practical detail, so essential to a decision in a particular context, has a softening and blurring effect when we are trying to think about what our priorities are in general.[251]

Glover also notes that fictional futures can also help us see the long-term implications of an approach which is viewed as acceptable, or unacceptable, from a short-term perspective.[252]

There are limitations, however, on the use of science fiction as a framework for philosophical and legal analysis. For example, speculation about the future can be wrong, particularly in terms of important details. In addition, as with any imaginative literature or film, science fiction places great value on good story telling, and the desire for raising and maintaining reader interest may trump concerns for clear, credible scientific or philosophical foundations or analysis. For example, though human slavery is disapproved by the "good guys,"[253] no human in the *Star Wars* movies shows any

---

251. GLOVER, *supra* note 146, at 17. For examples of similar views, see PARFIT, *supra* note 204, at 200 ("[Imagined] cases arouse in most of us strong beliefs. And these are beliefs, not about our words, but about ourselves. . . . Though our beliefs are revealed most clearly when we consider imaginary cases, these beliefs also cover actual cases, and our own lives."); SCIENCE FICTION AND PHILOSOPHY, *supra* note 179, at 1–2 (discussing importance of "thought experiments" to philosophy and arguing that "some of the best science fiction tales are in fact long versions of philosophical thought experiments"); ASHER SEIDEL, INHUMAN THOUGHTS: PHILOSOPHICAL EXPLORATIONS OF POSTHUMANITY 2–8 (2008) (discussing utility of philosophical analysis of speculative futures). An article on the movie *Avatar* notes:

> [G]ood science fiction operates on an allegorical level. In novels like "Dune," films like "Star Wars" or television series like the recent "Battlestar Galactica," . . . the fantastical elements of these works offer a place of "narrative safety" to contemplate real-life issues like environmental decay, totalitarianism and torture.
>
> "There's something very satisfying about being able to think through those issues without feeling you're actually taking a political position . . . . Because you're not—you're just talking about stories."

Dave Itzkoff, *You Saw What in 'Avatar'? Pass Those Glasses!*, N.Y. TIMES, Jan. 20, 2010, at A1 (quoting Annalee Newitz, Editor-in-Chief of the science fiction web site io9.com).

252. *See* GLOVER, *supra* note 146, at 17 (discussing how imagining possible futures can help shape current realities). Glover notes:

> [I]f we consider 'moderate' rather than extreme cases, our judgements are often influenced by awareness of being on a slippery slope. Rather than the extreme development being fully ignored, it lurks unconfronted as the thought of some nameless horror further down the slope. Another feature of thinking only about what at the time seems likely, is that we are hardly ever able to choose between kinds of life in general. We are limited to piecemeal decisions. A series of incremental decisions can lead us somewhere we would never have chosen to go in the first place, and rule out places we might have done well to go to.

Id.

253. *See, e.g.*, STAR WARS EPISODE I: THE PHANTOM MENACE (Lucasfilms 1999) (two of the film's central characters, Anakin Skywalker and his mother, Shmi, are slaves of a junk dealer, an exploitive relationship disapproved by Jedi Knight Obi-Wan Kenobi).

moral concern that sentient robots are treated as things[254] or that the first storm troopers, who were human clones designed and raised as property, are used to serve the Republic—i.e., the "good guys."[255] On the other hand, the *Star Wars* universe, like many science fiction stories, may be meant as a cautionary tale of the unjust and immoral things that could happen, rather than an endorsement of the practices involved.[256] Given the possibility of such a purpose, science fiction accounts should not be viewed simply as endorsements of particular practices. The purpose may be just the opposite. A final limit on using science fiction as a cultural lens is that it is too vast a genre for anything but a brief, illustrative discussion.

A. Modified Humans

The monster created by Frankenstein in Mary Shelley's novel is the most famous early science fiction portrayal of an enhanced or created human-like creature.[257] As with most accounts of artificial beings before the 1950s,[258] the science in *Frankenstein* is minimal. The scientific basis for the creation of Frankenstein's artificial man-like creature consists solely of considerable work assembling human parts scavenged from graveyards in a laboratory and an undefined method used to "infuse a spark of being into the lifeless" assemblage.[259] However, the work is a classic because of the power of

254. *See* DAVID WEST REYNOLDS, STAR WARS: THE VISUAL DICTIONARY 62 (1998) [hereinafter STAR WARS: DICTIONARY I] ("Droids are regarded as slaves and third-class citizens, held in contempt by those many who 'don't like machinery that talks back.'"). In order to hinder development in droids, their memories are regularly wiped. *Id.* at 7. Those which are not wiped tend "to develop identities and sentience." *Id.* Thus, some droids have "personalities and identities of their own." *Id.* at 62. An example of this class is R2-D2, who had "a mind of his own" and a "distinct personality and quirkiness." *Id.* at 20; DAVID WEST REYNOLDS, STAR WARS EPISODE I: VISUAL DICTIONARY 35 (1999) [hereinafter STAR WARS: DICTIONARY II]. Because of concerns about droids being either too independent or too lacking in intelligence, considerable effort went into controlling the abilities of droids that used weapons. STAR WARS: DICTIONARY I, *supra*, at 42 (explaining restrictions on droids' "abilities for independent action"); STAR WARS: DICTIONARY II, *supra*, at 21–23 (explaining droids were designed to be incapable of independent thought and have no ability to react to surprises or learn from experience).

255. ATTACK OF THE CLONES, *supra* note 162.

256. See *infra* notes 261, 280–82, and accompanying text for a discussion of the risk of hubris that can be involved in pursuing advances in technology and knowledge, and other cautionary "morality tales" in science fiction works. See *infra* note 276 and accompanying text for an example of a science fiction movie using the technique of role reversal to critique inequality.

257. MARY SHELLEY, FRANKENSTEIN (Brantley Johnson ed., Simon & Schuster Paperbacks 2004) (1818).

258. *See* THE VISUAL ENCYCLOPEDIA OF SCIENCE FICTION 172–73 (Brian Ash ed., 1977) [hereinafter ENCYCLOPEDIA OF SCIENCE FICTION] (describing development of functional robots in science fiction). For an example of the extreme lack of sophistication in depicting artificial humanlike creatures during the first part of the twentieth century, see EDGAR RICE BURROUGHS, SYNTHETIC MEN OF MARS 33–37 (Ballantine Books 1963). This story, which was originally published as a six-part serial in 1939 and published as a book in 1940, is the ninth book in the Barsoom (Mars) series. David A. Adams, *Synthetic Men of Mars, Summarized*, *in* EDGAR RICE BURROUGHS SUMMARY PROJECT (David Bruce Bogarth ed., 2002), *available at* http:www.erblist.com/erblist/syntheticsum.html. Burroughs's creatures are somewhat similar to the androids in Čapek's *R.U.R.*, which are discussed *infra* at notes 291–94, 298 and accompanying text.

259. SHELLEY, *supra* note 257, at 46–52, 55. The monster is treated herein as a modified human since human parts are used. It could also be viewed as an android. See *infra* notes 296–97 and accompanying text for a discussion of how an android may be defined.

its story and because it presents several themes that have become central to such stories. First, though the science in *Frankenstein* is sparse, the creature, who is never given a name, is depicted as the result of technology rather than magic or miracle.[260] Second, Frankenstein's actions highlight the risks of technology by showing "how dangerous is the acquirement of knowledge, and how much happier that man is who believes his native town to be the world, than he who aspires to become greater than his nature will allow."[261] Third, the creature's relationship with humans is basic to the story; because humans reject him rather than accept and nurture him, he develops and acts on an animosity toward humans.[262] Fourth, because of loneliness and a desire for offspring, the creature seeks a female like himself.[263] Fifth, Frankenstein refuses to provide a mate for the creature because of concerns that the creatures will replace humans as the dominant species.[264]

More recent science fiction depicts modified humans in a more scientific manner and deals with human modification in terms of genetic engineering, cloning, and cyborgs with mechanical or electronic enhancements.[265] Even though modified humans would seem to satisfy the requirements for personhood, their treatment in terms of normative personhood varies enormously. In large part, normative personhood depends on the reason for and the nature of the modification. For example, where genetic modifications are used to improve the abilities of humans, personhood is likely,[266] though discrimination and degrees of citizenship are possible.[267] On the other hand, genetic modifications to make a human more servile could be used to limit or deny normative personhood.[268] Stories based upon cloning vary widely, particularly in terms of the purpose for cloning. Many stories involve purposes that are potentially antithetical to normative personhood. Such purposes include mass production of clones

260. Scientific discoveries in Shelley's time involving the electrical natures of lightning and the apparent role of electricity in animating animals provided a scientific basis for the story. *See* PERKOWITZ, *supra* note 62, at 22–23.

261. SHELLEY, *supra* note 257, at 50. Frankenstein also notes that, even in the pursuit of knowledge, one should never "allow passion or a transitory desire to disturb his tranquility." *Id.* at 53. Frankenstein claims that, had this been done, many historical tragedies would have been avoided. *Id.* at 53–54. See *infra* note 282 and accompanying text for other discussions of hubris on the part of scientists.

262. SHELLEY, *supra* note 257, at 56–57, 153–71, 273.

263. *Id.* at 174–77.

264. *Id.* at 203–06.

265. For a useful review of legal issues implicated in science fiction accounts of cloning and genetic manipulation, see Christine Corcos et al., *Double-Take: A Second Look at Cloning, Science Fiction and Law*, 59 LA. L. REV. 1041 (1999).

266. *See* SLONCZEWSKI, *supra* note 116 (describing fictional ocean planet where genetic modifications produce physical changes (webbing in feet and hands) in a sea-based group ("Sharers") who share planet in mutual personhood with another group ("Elysians"), who have been modified to live over a thousand years).

267. *See, e.g.*, CHRIS MORIARTY, SPIN STATE (2003); CHRIS MORIARTY, SPIN CONTROL (2006). In Moriarty's future world, genetic "constructs" can be citizens of the "United Nations" but are subject to legal restrictions and to social discrimination. See *infra* notes 270, 355–64 and accompanying text for further discussion of these books. *See also* GATTACA (Columbia Pictures 1997) (presenting future world with caste society in which those with access to genetic engineering compose a superior class).

268. *See, e.g.*, MAUREEN F. MCHUGH, NEKROPOLIS 36–40 (2001) (describing world in which some, but not all, countries allow: (1) people to become "jessed" (a technology that renders one subservient to an "owner"), and (2) human-like biological constructs ("harnis") designed to be loyal and compliant).

in order to meet a need for large quantities of one type of human for specific purposes, such as providing soldiers for an army[269] or workers for space environments.[270] An individual clone of a specific person to provide a source of organs clearly involves a denial of the clone's personhood.[271]

Stories about cyborgs vary greatly, both in terms of the purpose and degree of enhancement. Where the enhancement is subtle, legal and social personhood are generally granted, just as minor mechanical implants like cochlear implants and hip replacements are treated today.[272] Where the enhancement results in humans that are more visually machine-like, negative bias and discrimination exist even if autonomous personhood has been granted.[273] It is hard to know how to categorize humans who have transferred their memories and self-consciousness to digital form. Are these copies modified humans or a form of machine entity? Regardless of the proper classification, where they are placed in a "body" of some sort, they are generally treated as persons.[274]

269. See *supra* note 255 and accompanying text for a discussion of the use of clones as soldiers in *Attack of the Clones*.

270. In SPIN CONTROL, *supra* note 267, and SPIN STATE, *supra* note 267, cloned genetic constructs, raised in crèches with thousands of clones like one another, are viewed as property (with the possibility of citizenship after 30 years) to be used by "the syndicates," a society living largely on space stations. See *supra* note 267 and *infra* notes 355–64 and accompanying text for further discussion of this future world. In C.J. CHERRYH, CYTEEN (1988), cloning and genetic manipulation provide the population needed for expansion. Though most clones are property to be used as servants or soldiers, the replacement clone of a citizen is treated as a "replicate" citizen. *Id.* at 52.

271. *See, e.g.*, THE ISLAND (DreamWorks SKG 2005) (depicting illegal facility for producing and maintaining clones to provide organs for people who have cloned themselves).

272. For example, in HARRISON & MINSKY, *supra* note 151, at 100–01, 128–29, 373–75, a man, who has had processors inserted in his injured brain to enable him to communicate directly with a self-conscious machine intelligence system, becomes more machine-like in his behavior and attitudes. However, he is a full legal person and is not subject to discrimination.

Humans with enhancements for direct human-machine interface are common in science fiction and are generally treated as persons. *See, e.g.*, WALTER JON WILLIAMS, HARDWIRED 50–51 (1986) (depicting a man with implanted sockets that enable him to connect directly with circuitry of armored tank).

273. *See, e.g.*, MCCAFFREY, THE SHIP WHO SANG, *supra* note 123 (portraying frequent human misunderstanding of and prejudice toward humans who, as hopelessly malformed children, had been fitted into spaceships, which they operate and which serve as their mechanical bodies); FREDERIK POHL, MAN PLUS 27–29 (1976) (portraying negative reactions by most humans to a cybernetic human capable of living on Mars).

274. For example, three novels by Richard K. Morgan—WOKEN FURIES (2005), BROKEN ANGELS (2003), and ALTERED CARBON 12–13 (2002)—take place about six centuries in the future in a world where people can be digitally copied and stored in a machine or downloaded into a small storage device (a "stack") and inserted into a body (a "sleeve"), which can be, but need not be, a clone of the person's own body. A similar treatment of clones occurs in WALTER JON WILLIAMS, VOICE OF THE WHIRLWIND (1987), in which stored copies of a person's mind are placed in a clone of that person when he dies. Where a digital copy of a human is installed in a machine, it tends to be treated as a person. *See, e.g.*, ROGER MACBRIDE ALLEN, THE MODULAR MAN 19 (1992) (man whose "self" has been transferred into robotic vacuum cleaner with enhanced capacities is treated as person by those who were close to him prior to transferral and, following trial, is legally recognized as human); CHARLES PLATT, THE SILICON MAN 248–53 (1991) (describing a future where downloads live together in a large simulation and have rights of humans, enjoy ability to interact with physical world, and provide services through terminals and robotic "bodies"); SAWYER, *supra* note 241, at 162–63, 345, 354–59 (though the result of litigation over personhood for mechanical androids with downloaded human minds is not resolved, epilogue suggests that downloaded entities are persons). For a discussion of the possibility for such downloaded (or uploaded) copies, see *supra* notes 66–68 and accompanying text.

Even where they are totally digital, there is also a tendency to grant them personhood.[275] This choice probably results more from the necessities of storytelling than from a view of justice and morality.

B. Modified Animals

The intellectual enhancement of animals has received less attention than the intellectual enhancement of humans or machines. Apart from scenarios like that in the movie *Planet of the Apes*,[276] which provides perspective by reversing the respective roles of apes and humans, most science fiction depicts enhanced animals as property to be used as soldiers,[277] slaves,[278] a source for spare organs, or test subjects for experiments.[279] Some of the stories are, in effect, morality tales[280] of the potential disastrous results of scientific developments[281] and of the hubris or evil of at least some

275. Greg Egan has explored this concept in two novels, DIASPORA (1998) and PERMUTATION CITY (1994), which sketch a fully realized world of totally digital persons. *See also* DENNIS DANVERS, END OF DAYS (1999); DENNIS DANVERS, CIRCUIT OF HEAVEN (1998) (describing near-future world where nearly all humans have downloaded themselves into in a virtual utopia).

276. PLANET OF THE APES (APJAC Productions 1968). In this movie, human astronauts crash on a planet where various species of apes with speech and technology oppress and enslave the "primitive" humans.

277. *See, e.g.*, S. ANDREW SWANN, MOREAU OMNIBUS (2003). This collection of three novels by Swann presents a near future where enhanced animals were developed to serve as soldiers in a series of wars. These enhanced animals are called Moreaus, after the creatures in H.G. Wells' novel, THE ISLAND OF DOCTOR MOREAU (1896). At the end of the wars, the surviving Moreaus suffer from severe speciesist discrimination despite eventual recognition of their claim of constitutional rights as persons by the Supreme Court. SWANN, *supra*, app. at 700–01. Wells's novel depicts an isolated island where Dr. Moreau, acting out of a combination of scientific curiosity and hubris, has created enhanced beasts called Beast Men, who are subhuman constructs able to speak and reason crudely. *See* WELLS, *supra*.

278. *See, e.g.*, STEPHEN BAXTER, MANIFOLD TIME (2000) (genetically modified squid used to pilot space craft); DI FILIPPO, *supra* note 170 (describing future where human-animal "splices" are common and where only splices with over 50% human genetic make-up are treated as persons while those with a lesser percentage are, in effect, slaves); SLONCZEWSKI, *supra* note 116, at 56 (describing far-future world where human-gorilla hybrids are bred to serve as slaves); CORDWAINER SMITH, NORSTRILIA 8 (1988) (describing genetically enhanced animals who are enslaved). In a prequel to *Planet of the Apes*, humans use intelligent, nonverbal higher apes as enslaved servants. CONQUEST OF THE PLANET OF THE APES (APJAC Productions 1972).

279. *See, e.g.*, STEPHEN GALLAGHER, CHIMERA 262–65 (1982) (depicting experiments with benefits that could include provision of spare organs); OLAF STAPLEDON, SIRIUS 20–22 (1944) (portraying experimental increase in canine intelligence); WELLS, *supra* note 277 (depicting experimental enhancement of animals).

280. *See, e.g.*, DAVID BRIN, HEAVEN'S REACH *Afterword* 431 (1998). In the afterword to this novel concerning enhanced animals, Brin notes that he deliberately took a more positive note:

> I also noticed that nearly all these [earlier] tales [of genetically altered nonsapient animals] assume that human "masters" will always do the maximally stupid/evil thing. In other words, if we meddle with animals to raise their intelligence, it will be in order to enslave and abuse them.
>
> . . . I feel it is now unlikely our civilization would behave in a deliberately vile way toward newly sapient creatures, because the morality tales did their job!

Id. at 431. For further discussion of this novel and related novels, see *infra* notes 285–88 and accompanying text.

281. *See, e.g.*, DEAN KOONTZ, SEIZE THE NIGHT 307 (paperback ed. 1999) (depicting "doomsday" scenario caused by scientist's "misplaced trust in the power of science to resolve all problems and explain all things").

scientists.[282] Even where enhanced animals are legally free, it is common to depict them as members of an underclass, subject to exploitation and discrimination.[283]

There are, however, some positive presentations of the role of enhanced animals. For example, some of the enslaved animals eventually gain freedom.[284] David Brin has constructed an "Uplift Universe" where advanced species (patrons) uplift—i.e., genetically enhance—less advanced species (clients) and then hold extensive power over their client species.[285] Humans, in contrast, have uplifted dolphins and chimpanzees while also granting them civil rights.[286] Even where animals lack rights or face discrimination, some humans show respect and affection for them as persons.[287] Nevertheless, where an enhanced animal is treated as a person, there can still be problems caused by a lack of hands or human senses and from a lack of the companionship of similarly enhanced members of its species.[288]

C. Machines

Science fiction did not address machines that possess substantial intelligence until the twentieth century.[289] Beginning in the 1920s, robots became a popular subject.[290] The term "robots" originated in a science fiction play, *R.U.R. (Rossum's Universal Robots)*,[291] by Karel Čapek, which premiered in 1921.[292] Though the artificial entities

282. For example, *Chimera* and *The Island of Dr. Moreau* both involve sadistic megalomaniacal scientists. *See* GALLAGHER, *supra* note 279; WELLS, *supra* note 277. *Seize the Night* includes scientists causing harm by either evil or hubris. *See* KOONTZ, *supra* note 281.

283. *See, e.g.*, SWANN, *supra* note 277, at 700–01 (describing treatment of enhanced animals, known as moreaus, as refugees with anti-Moreau sentiment prevalent in nation).

284. *See, e.g.*, BAXTER, *supra* note 278, at 439–40 (illustrating scene where descendants of genetically modified squids leave solar system); SLONCZEWSKI, *supra* note 116, at 410–11 (describing legal rights earned by gorilla-human hybrid slaves after struggle for freedom); SMITH, *supra* note 278, at 245–48 (portraying rebel group of genetically modified animals, called "underpeople," who escape enslavement and live underground with their own government); SWANN, *supra* note 277, at 698 (presenting scenario where United States grants genetically enhanced animals protections under the Bill of Rights).

285. *See, e.g.*, BRIN, *supra* note 280, at 447 (defining "Uplift" process). There are six novels in the series.

286. *Id.* at 445.

287. *See, e.g.*, STAPLEDON, *supra* note 279, at 155–60 (depicting story of human female who treats enhanced dog, with whom she was raised, as a brother and (perhaps) as a lover); SWANN, *supra* note 277, at 130–35 (telling story of human female who forms liaison with and eventually marries enhanced tiger).

288. For example, the owners and "family" of the enhanced dog in Stapledon's SIRIUS, *supra* note 279, at 166–67, treat the dog as a person in many ways; but the dog, lacking hands, cannot write or use tools well and, as the only such enhanced dog, is unable to share his superior natural ability to enjoy smells and sounds.

289. *See* ENCYCLOPEDIA OF SCIENCE FICTION, *supra* note 258, at 172–73 (describing development of functional robots in science fiction); For discussion of fantasy treatments of mechanical "persons," see PERKOWITZ, *supra* note 62, at 17–21, discussing fantasy treatments of "mechanical persons."

290. *See* ENCYCLOPEDIA OF SCIENCE FICTION, *supra* note 258, at 173 (describing rise of robots in science fiction); *see also* PERKOWITZ, *supra* note 62, at 21–38 (tracing history of robots in science fiction).

291. KAREL ČAPEK, *R.U.R. (Rossum's Universal Robots)*, *reprinted in* TOWARD THE RADICAL CENTER: A KAREL ČAPEK READER 34 (Peter Kussi ed., 1990).

292. *Chronology*, *in* TOWARD THE RADICAL CENTER, *supra* note 291, at 28. In its original Czeck version, the play used the term "robota," which means "heavy labor." ČAPEK, *supra* note 291, at 33.

in the play are more aptly viewed as organic artifacts,[293] the term "robot" has come to refer to machines.[294] Robots can take many forms, but humanoid robots are the most popular form in science fiction.[295] Such robots are sometimes termed "androids,"[296] but some writers restrict the term android to humanoid robots with synthetic biological or chemical components that are grown rather than a humanoid mechanical entity that is manufactured.[297] The robots in Čapek's play were such synthesized androids.[298] To the extent that organic androids involve the manipulation of human DNA, they overlap with the category of modified humans. Another area of overlap involves "downloads" of human minds to an electronic form. Where these downloaded minds continue to interact with the physical world, whether in a borrowed human body or a mobile robotic form, they tend to be treated as human and thus as persons.[299] Because of their more recent development and lesser anthropomorphic appeal, computers have received less attention in science fiction than robots and androids,[300] even though computers may be the first machines with the computational power necessary for consciousness. Because of this possibility, more recent science fiction shows an increased consideration of supercomputers as examples of autonomous or semi-autonomous artifacts.[301]

Depictions of robots and computer entities are so numerous and diverse that there is no way to review them herein in detail. Instead, this Article will address five different visions of the relationship between humans and self-conscious machines: (1) war and human subjugation, (2) enslavement of machinery, (3) humans and machines in a modus vivendi, (4) humans and machines in a shared society based on respect and equality, and (5) transcendent supermachines that exist largely apart from human societies. There are also stories that address preventing the development of self-

293. *See* ČAPEK, *supra* note 291, at 38–42 (depicting artificial entities grown from organic living matter, engineered, and redesigned for mass production). The movie *Metropolis* also featured a robot, which was referred to in the movie as a "machine man," though it was actually a female humanoid robot. METROPOLIS (Universum Film AG (UFA) 1927).

294. ENCYCLOPEDIA OF SCIENCE FICTION, *supra* note 258, at 172 (explaining that robots may be defined as entities, often made of metal, whose minds are mechanical devices). For a discussion of the development of the term "robot," see Jana Horáková & Josef Kelemen, *The Robot Story: Why Robots Were Born and How They Grew Up*, *in* THE MECHANICAL MIND IN HISTORY 283–306 (Philip Husbands et al. eds., 2008).

295. *See* ENCYCLOPEDIA OF SCIENCE FICTION, *supra* note 258, at 175–80 (discussing humanoid robots).

296. *Id*. at 180.

297. *Id*. at 161, 180. Isaac Asimov adopts the following distinction between robots and androids:

> A robot may be defined as a mobile artefact, made of metal, that can usually think for itself. It may or may not look like a human being . . . .
>
> . . . .
>
> Androids may be defined as 'robots made of flesh'. While they can be programmed to accept orders in the same way as robots, their bodies are chemically or biologically based and are grown rather than built.

Id. at 172 (quoting Isaac Asimov).

298. See *supra* note 293 for a discussion of the androids in Čapek's play.

299. See *supra* note 274 and accompanying text for a discussion of downloads.

300. *Compare* ENCYCLOPEDIA OF SCIENCE FICTION, *supra* note 258, at 171–80 (discussing computers), *with id.* at 181–84 (discussing cybernetics).

301. *See* 2001: A SPACE ODYSSEY (MGM et al. 1968) (depicting self-conscious computer). See *infra* notes 303, 305–06, 311, 366–80, and accompanying text for additional examples of self-conscious computers.

conscious machines or terminating existing ones. However, perhaps because of better possibilities in telling dramatic stories, many of these involve at least a partial failure in the attempt resulting from the difficulty of prevention,[302] the desire to enjoy the benefits of highly intelligent machines,[303] or a continuing struggle in which some, but not all, intelligent machines have become self-conscious.[304]

1. War and Human Subjugation

The Terminator series,[305] *The Matrix* series,[306] and the *Battlestar Galactica* series[307] all take place in a future world in which self-conscious machines are at war with humans. All three of these series present a bleak world in which embattled humans have been defeated or subjugated and struggle for survival. Obviously, such a future is undesirable to humans in terms of both prudence and morality.

2. Beneficent Mechanical Slaves

It is virtually impossible to write about science fiction and robots without discussing Isaac Asimov, who addressed the subject of robots and computers in several novels,[308] nearly forty short stories,[309] and numerous nonfiction articles.[310] Because the

302. *See* GIBSON, *supra* note 170 (depicting two AIs succeeding in breaking free of controls); Thomson, *supra* note 40 (female android ("Maggie") built in violation of anti-AI laws by socially awkward man who desired female companionship develops self-consciousness and secretly achieves independence for herself and other AIs).

303. *See, e.g.*, LARRY NIVEN & BRENDA COOPER, BUILDING HARLEQUIN'S MOON (2005) (describing humans, fleeing an earth where machines and nano have escaped control, reluctantly keeping highly intelligent AI because they need it to fly their spaceship and to establish a new world).

304. See *supra* note 254 for a discussion of efforts in the *Star Wars* universe to prevent mechanical "droids" from developing consciousness. *See also* SLONCZEWSKI, *supra* note 116, at 329 (depicting human owners working to prevent self-consciousness in their AIs after self-conscious AIs were granted freedom).

305. THE TERMINATOR (Hemdale Film Corp. et al. 1984); TERMINATOR 2: JUDGMENT DAY (Carolco Pictures et al. 1991); TERMINATOR 3: RISE OF THE MACHINES (C-2 Pictures et al. 2003); TERMINATOR SALVATION (The Halcyon Company et al. 2009). In this series, Skynet is an AI system that controlled a computerized weapon system for the United States. When humans attempted to shut it down, Skynet, which had acquired sentience, launched a nuclear war and subsequently subjugated the human survivors and fought with a small group of human rebels.

306. THE MATRIX (Groucho II Film Partnership et al. 1999); THE MATRIX RELOADED (Warner Bros. Pictures et al. 2003); THE MATRIX REVOLUTIONS (Warner Bros. Pictures et al. 2003). In this series, the future earth is dominated by machines. Most humans are kept alive in pods where they experience a virtual reality in the "matrix," which makes them believe they live normal lives, even though in fact they are being used by the machines as a source of energy.

307. *Battlestar Galactica* was a television series that initially aired in 1978. *Battlestar Galactica* (Glen. A. Larson Productions & Universal TV 1978). The action takes place in another part of the galaxy where humans have been attacked by robotic Cylons, which they had created. The humans are forced to flee into space, chased by the Cylons, in search of a mythical home on a planet called Earth. In 2004, a new series based on the same basic plot premise aired. *Battlestar Galactica* (British Sky Broadcasting et al. 2004).

308. The novels are: THE CAVES OF STEEL (1954), THE NAKED SUN (Fawcett Crest 1972) (1957), THE ROBOTS OF DAWN (1983), and ROBOTS AND EMPIRE (Del Rey 1985).

309. JAMES GUNN, ISAAC ASIMOV: THE FOUNDATIONS OF SCIENCE FICTION 41 (Scarecrow Press rev. ed. 1996).

310. *See, e.g.*, ENCYCLOPEDIA OF SCIENCE FICTION, *supra* note 258, at 172 (nonfiction article by Asimov distinguishing robots and androids). See *supra* note 297 for a quote from this article.

short stories are very diverse,[311] the following discussion focuses on the novels that involve only robots and form a more coherent scheme. Asimov viewed "robots as industrial products built by matter-of-fact engineers," which "were built with safety features."[312] The primary safety feature is the built-in control based on the famous three laws of robotics:

1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[313]

These laws are so deeply engrained that robots "cannot imagine anything else" than being bound by them.[314]

A fourth, more basic law is developed and accepted by two intelligent humanoid robots "as a corollary of the First Law."[315] This "Zeroth Law" is: "A robot may not injure humanity or, through inaction, allow humanity to come to harm."[316] The robots independently "programmed" themselves to understand and be bound by this law.[317] Because the Zeroth Law was more basic than the others, it could compel a robot, in

311. Of particular interest to the issue of machine personhood are Asimov's short stories that explore other dimensions of life with intelligent machines. For example, there are stories indicating the following scenarios: (1) Machines become benevolent tyrants. *See, e.g.*, *The Evitable Conflict*, *in* ISAAC ASIMOV, THE COMPLETE ROBOT 447 (1982) [hereinafter COMPLETE ROBOT]; *The Life and Times of Multivac*, *in* ISAAC ASIMOV, THE BICENTENNIAL MAN AND OTHER STORIES 114 (1976). (2) Robots can pass as humans and assume political power. *See, e.g.*, *Evidence*, *in* COMPLETE ROBOT, *supra*, at 425; *The Tercentenary Incident*, *in* COMPLETE ROBOT, *supra*, at 187. (3) Robots replace humans. *See, e.g.*, *Robot Visions*, *in* ASIMOV, ROBOT VISIONS 19 (1990); *. . . That Thou Art Mindful of Him*, *in* COMPLETE ROBOT, *supra*, at 495. (4) Humans act to prevent robots from viewing themselves as autonomous. *See, e.g.*, *Robot Dreams*, *in* ISAAC ASIMOV, ROBOT DREAMS 25 (Ace Book 2d paperback ed. 2004). (5) As humans become more mechanical and robots become more organic, the line between them becomes blurred. *See, e.g.*, *Segregationist*, *in* COMPLETE ROBOT, *supra*, at 127.

312. COMPLETE ROBOT, *supra* note 311, at xii; *see also* ROBOT VISIONS, *supra* note 311, at 8 ("The stories were . . . portrayals of future technology . . . . The robots were machines . . . ."); *id.* at 453 ("I made them machines to serve human ends.").

313. ROBOT VISIONS, *supra* note 311, at 8; THE NAKED SUN, *supra* note 308, at 33.

314. ROBOTS AND EMPIRE, *supra* note 308, at 36 (internal quotation marks omitted).

315. *Id.* at 426 (internal quotation marks omitted).

316. *Id.* at 353 (internal quotation marks omitted).

317. *Id.* at 463. The concept of the Zeroeth Law had been stated to the robot (Danee Olivaw) by a human (Elijah Baley) in an earlier novel in the following exchange:

> It is as much my job to prevent harm to mankind as a whole as yours is to prevent harm to man as an individual. Do you see?
>
> I do not, Partner Elijah.
>
> Then that is because you're not made to see. Take my word for it that if you were a man, you would see.

THE NAKED SUN, *supra* note 308, at 108.

order to further the good of humanity, to harm humans, disobey humans, or show greater concern for a robot than humans.[318]

These "laws" are obviously far too brief and indeterminate to be applied properly unless the robots have the ability to exercise human levels of judgment and analysis. They are also unable to prevent robots from asserting their freedom or from harming humans.[319] For example, the laws are circumvented in one novel by redefining human beings as only those humans living on a particular planet and then ordering robots to kill humans from other planets.[320] The three laws also failed to prevent the freedom of thought reflected in the articulation of the Zeroth Law. Similarly, any machine with the ability to apply the three laws with any intellectual sophistication is very likely to have the ability to evade the laws by programming itself in a manner like that used to develop the Zeroth Law—for example, by defining itself as the functional equivalent of a human.[321] Finally, Asimov's robots also obey robots who look like humans.[322]

318. For an example of robots disobeying a human, see ROBOTS AND EMPIRE, *supra* note 308, at 355–56. In a short story, a robot, who is an exact duplicate of the President, uses a version of the Zeroeth law to justify the killing of the President because the robot would be able to be a better President and thus able to save many more lives. *The Tercentenary Incident*, *in* COMPLETE ROBOT, *supra* note 311, at 197–99. Similarly, a self-conscious robot in one of the robot novels tells another robot that the other robot is more important to humanity than a human and that therefore, "the Zeroeth Law demands that I protect you above all else." ROBOTS AND EMPIRE, *supra* note 308, at 426.

319. See *supra* notes 317–18 and *infra* notes 320–21, 343, and accompanying text. One difficulty with the Three Laws is that the prohibition of allowing "harm" to a human is similar to John Stuart Mill's "harm principle" that "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others." MILL, *supra* note 18, at 956. Defining "harm to others" has presented enormous difficulties in the application of this principle. *See, e.g.*, NORMAN E. BOWIE & ROBERT L. SIMON, THE INDIVIDUAL AND THE POLITICAL ORDER: INTRODUCTION TO SOCIAL AND POLITICAL PHILOSOPHY 112–20 (4th ed. 2008) (discussing difficulties with Mill's harm principle). The Zeroth Law is a crude form of utilitarianism, and thus subject to the problems involved with stating and applying utilitarianism. *See, e.g.*, KYMLICKA, *supra* note 8, at 10–48 (listing problems with utilitarianism, such as difficulty in defining utility and its maximization, and the inadequate conception of equality); BOWIE & SIMON, *supra*, at 29–48 (discussing and critiquing utilitarianism). For criticism of the practicality of programming machines in accordance with Asimov's laws, see BROOKS, *supra* note 38, at 203–04 (stressing limited nature of robotic intelligence and complexity of the laws); SINGER, *supra* note 156, at 422–23 (noting that the three laws are a fictional plot device and that they cannot be programmed). Another problem with relying on Asimov's laws is that so much current research and development is directed toward improving robotic weapons, which clearly do not satisfy the laws. *See* BROOKS, *supra* note 38, at 204 (suggesting prohibition on use of robotic weapons as way to address limits of human control); SINGER, *supra* note 156, at 423, 435–36 (noting that military robots do not follow the laws and that military officials refuse to acknowledge the societal impact of developing robotics for warfare).

320. ROBOTS AND EMPIRE, *supra* note 308, at 154–57. The difficulty, in part, is that while robots cannot knowingly kill a human, they may have inadequate knowledge and, as a result, kill humans. *See* THE NAKED SUN, *supra* note 308, at 163–66 (discussion between characters where one makes this broad point); THE CAVES OF STEEL, *supra* note 308, at 184–85 (robot unknowingly assisted in an accidental homicide by delivering the weapon used to commit it).

321. *See . . . That Thou Art Mindful of Him*, *in* COMPLETE ROBOT, *supra* note 311, at 517 (two robots conclude they are human and that their kind should replace humans).

322. *See* THE NAKED SUN, *supra* note 308, at 106–07 (portraying human informing robots that robot they previously viewed as master was actually a robot).

These problems with the laws are largely irrelevant because their purpose was not to provide effective control of robots. Asimov was concerned instead with two other goals. First, he wanted an alternative to the two traditional presentations of robots:

> In the first class there was Robot-as-Menace. I don't have to explain that overmuch. Such stories were a mixture of "clank-clank" and "aarghh" and "There are some things man was not meant to know." After a while, they palled dreadfully and I couldn't stand them.
>
> In the second class (a much smaller one) there was Robot-as-Pathos. In such stories the robots were lovable and were usually put upon by cruel human beings.[323]

The second, and perhaps more important goal, was to provide puzzles to be solved.[324] For example, the puzzle of robots which were able to kill humans, despite the three laws, was solved by discovering that, for these robots, humans had been redefined in a manner that excluded some humans.[325]

To the extent that the three laws are effective, they function as a self-executing slave code for self-conscious robots: Do not harm masters; obey masters except where harm to a master would result; and protect your owner's property interest in your well-being.[326] In one of the novels, robots are explicitly compared with the enslaved Helots in Sparta.[327] In contrast to the Helots' ability to revolt, the three laws prevent the mechanical slaves from doing so.[328] Nevertheless, some owners of the robots had a fear that some humans might teach the robots to circumvent the laws and become "capable of revolting."[329]

There are two societies of humans in Asimov's robot novels. One does not use robots because they believe "that a robot society is bound to decay" due to the ease and convenience of living with robotic servants.[330] The other society embraces the use of robots because, as mechanical slaves, the robots enable their owners to enjoy high productivity and considerable leisure. As personal servants, they are extremely desirable. Gladia Delmarra, a central figure in the robot novels, expresses her views as follows: "[W]hat pleasure could be greater than that of being taken care of by your own robots, robots who knew your every signal, and, for that matter, knew your ways and

323. COMPLETE ROBOT, *supra* note 311, at xi.

324. *See, e.g.*, GUNN, *supra* note 309, at 53 ("The robots exist to present the puzzle in their behavior; the characters exist to solve the puzzle.").

325. See *supra* note 320 and accompanying text for this example of robots redefining humans in Asimov's work. For an example of robots redefining themselves as human, see *supra* note 321 and accompanying text.

326. Asimov was aware of this effect. In *The Bicentennial Man*, a robot who requests freedom in a judicial proceeding is granted its request based on the following reasoning by the judge: "There is no right to deny freedom to any object with a mind advanced enough to grasp the concept and desire the state." ASIMOV, COMPLETE ROBOT, *supra* note 311, at 519, 528.

327. NAKED SUN, *supra* note 308, at 121–22.

328. *Id.* at 122. For an example of the use of a drug to achieve the same result in a society with human slaves, see ELIZABETH A. LYNN, THE SARDONYX NET 18 (Berkley ed. 1982).

329. NAKED SUN, *supra* note 308, at 210.

330. ROBOTS AND EMPIRE, *supra* note 308, at 97.

desires even without being told."[331] Given this pleasure, it is not surprising that one character enjoyed having a humanoid robot live with her as a "Stepford husband."[332]

Despite their slave status, some humanoid robots are not simply machines, but have self-consciousness, a strong sense of identity, and considerable intelligence and creativity. Daneel Olivaw, for example, was a robot who worked closely with a human named Elijah Baley. Both have major roles in the novels. Their interaction shows that Daneel clearly passes the functional test adopted in this Article.[333] This ability creates a tension within Baley. On the one hand, Baley tells himself, "[t]he thing was only a machine."[334] On the other hand, Baley clearly regards Daneel as a human equivalent and "friend," not a robot.[335] Baley tells Daneel to "[s]top insisting" that he is a robot because, "[y]ou could not mean more to me, Daneel, if you were a man."[336] A similar tension exists in Gladia Delmarra. In the final robot novel, she delivers an impassioned and successful speech in favor of the rights of intelligent humanoid robots like Daneel and a similar humanoid robot, Giskard: "It is not enough to have respect for all human beings; one must have respect for all intelligent beings."[337] However, within hours of the speech, she knowingly uses Giskard as a menial servant who has to do her bidding.[338] In Asimov's robot world, humans with mechanical slaves are simply unable to grant freedom to these useful minions.[339]

Ironically, Asimov's novels present robots both as menace and as pathos, despite his desire to find an alternative model.[340] The pathos image is reflected in Daneel and Giskard, who are sufficiently intelligent and self-conscious to see themselves as the same as humans but, nonetheless, feel compelled to follow the laws of robotics and accept their treatment as things by humans.[341] The robot-as-monster image manifests

331. *Id.* at 106.

332. *See* ROBOTS OF DAWN, *supra* note 308, at 135–41. This character was a resident of the planet Aurora, which had manufactured the humanoid robot. One Auroran summarized the Auroran perspective on the use of humanoid robots as sex partners:

> Aurorans had no objection to such a situation in theory. When they stopped to think of it, however, Auroran women simply did not enjoy the thought of having to compete with robot women. Nor did Auroran men wish to compete with robot men.

ROBOTS AND EMPIRE, *supra* note 308, at 276. Partly for this reason, the production of additional humanoid robots was abandoned. *Id.* See also *supra* note 40 and accompanying text for a discussion of "Stepford wives" and other science fiction treatment of romantic and sexual relationships with humanoid robots.

333. Asimov noted that Daneel and Giskard, another humanoid robot, "were advanced to the point where their minds were of human complexity." ROBOT VISIONS, *supra* note 311, at 481.

334. THE NAKED SUN, *supra* note 308, at 189.

335. ROBOTS AND EMPIRE, *supra* note 308, at 226; THE NAKED SUN, *supra* note 308, at 24.

336. ROBOTS AND EMPIRE, *supra* note 308, at 227.

337. *Id.* at 214.

338. *Id.* at 215.

339. See *supra* notes 114, 330–32, and accompanying text for a discussion of human enjoyment of and reliance on slaves and the consequent reluctance to free them.

340. See *supra* note 323 and accompanying text for a discussion of Asimov's desire to move beyond the traditional presentations of robots.

341. *See* ROBOTS AND EMPIRE, *supra* note 308, at 362–68. Similar pathos is involved in one of Asimov's short stories in which a robot's perspective is so warped by his "inferior status" that he chooses to die so that he can be accepted as a human. *See The Bicentennial Man*, *in* COMPLETE ROBOT, *supra* note 311, at 519, 556–

itself in several ways. First, the benefits of robots cause their owners' societies to trade leisure and convenience for dynamism and growth, leading to the erosion of the societies' moral fabric.[342] Second, the menace of robots unconstrained by the limits of the three laws is shown by the ease of redefining human to exclude most humans. Finally, perhaps the greatest menace is shown in Giskard's decision to satisfy the Zeroth Law by allowing Earth to become radioactive and uninhabitable so that humanity will be forced, for its own good, to go to the stars.[343] Such a "sorcerer's apprentice" use of power without adequate wisdom is a powerful cautionary tale. Asimov's short stories also indicate that he was aware that robots might not be benevolent and that benevolent intelligent machines could cause unexpected undesirable consequences. For example, several of his short stories concern machines that, because they are machines bound by the "Zeroth Law,"[344] decide that only the machines can best know "what the ultimate good of Humanity will entail."[345] The machines use their power to pursue this goal, even if harm to some humans results, and refuse to explain to humans what is happening because that "may make . . . [humans] unhappy and may hurt . . . [human] pride."[346]

3. Modus Vivendi

Rudy Rucker's robot novels are, to a considerable extent, the opposite of Asimov's in both tone and content.[347] Unlike the safety-conscious "matter-of-fact" engineers who create safe, well-designed robots in Asimov's novels, Cobb Anderson, the creator of the self-conscious robots in Rucker's novels, deliberately enabled the robots to program themselves so that they could evade the "Three Laws of Robotics."[348] Anderson *wanted* them to revolt; he "didn't want to father a race of slaves."[349] (He was tried for treason for this, but was acquitted.)[350] Even Anderson's lifestyle is "anti-Asimovian." At the start of the first book, Anderson is a retired "pheezer" (freaky geezer), who lives in South Florida and has a fondness for cheap

57. This Pinocchio-like desire to be a "real" human is common in tales of androids. *See, e.g.*, PERKOWITZ, *supra* note 62, at 47–48 (discussing *Star Trek* and *A.I. Artificial Intelligence*).

342. The desire for growth and dynamism erodes the moral fabric of society as it results in people like Gladia Delmarra deciding to treat the robot Giskard as a slave even though she knew it was wrong to deny him the respect she would show a human. See *supra* notes 337–38 and accompanying text. *See also* THE NAKED SUN, *supra* note 308, at 127–28 (noting decline of size of population, static quality, and focus on happiness of planets with robots).

343. ROBOTS AND EMPIRE, *supra* note 308, at 464–67.

344. See *supra* notes 313–29 and accompanying text for a discussion of the fundamental laws of robotics.

345. *See, e.g.*, *The Evitable Conflict*, *in* COMPLETE ROBOT *supra* note 311, at 447, 468–69 (describing robots who violate the three laws and attempt to control mankind).

346. *Id.* at 468.

347. In chronological order, the four novels are SOFTWARE (Avon 1987) (1982), WETWARE (1988), FREEWARE (1997), and REALWARE (2000). Perhaps the ultimate anti-Asimovian robot is Bender in the television series *Futurama* (20th Century Fox Television, 1999–2010).

348. SOFTWARE, *supra* note 347, at 79.

349. *Id.*

350. *Id.* at 80.

sherry.[351] This is a far cry from the restrained, dependable 1950s characters like Elijah Baley in Asimov's stories.

Anderson's robots developed consciousness in a "survival of the fittest" process characterized by intense competition among themselves and by random mutations and mergers.[352] The robotic society that emerged is an anarchic modus vivendi in which each robot behaves in the manner of the rational selfish person underlying economic theory, whether in dealing with another robot or with humans.[353] Though most humans in Rucker's novels feel considerable animosity toward the robots, there is no way to subjugate the robots because the robots have established a secure haven on the moon. In addition, the humans benefit through trade with the robots. For example, the robots supply humans with drugs manufactured in the sterile, low temperature conditions on the moon and with replacement human organs that have been grown on the moon.[354]

A similar modus vivendi arrangement is presented in two novels by Chris Moriarty set in a future interstellar human civilization shared by "true" humans (almost all of whom are posthumans), genetically constructed "humans," and AI systems.[355] The AIs vary enormously in size, abilities, and attitudes toward humans. The larger, more capable systems are self-aware, have achieved autonomy, and live in modus vivendi relationships with humans and each other. Humans and AIs can communicate with one another through the "spinstream," which is akin to a large, instantaneous interstellar internet.[356] In addition, an AI can use a "shunt" implanted in a human that enables it to operate the human's body and thereby interact physically with other humans.[357]

One of the most interesting characters in the novels is an "emotive" AI, named Cohen, with enormous power over spinstream communications.[358] Though he utilizes machines and humans (by the use of shunts) to interact physically with the world, Cohen is basically an extraordinarily powerful, complex, self-conscious digital program.[359] Though he values his survival and freedom highly,[360] Cohen provides a model of what could be viewed as a partially Asimovian AI. A core program in Cohen is the "Game," which directs him to track and maximize "player hits," defined as positive emotive cues in terms of increases in length and intensity of play and explicit player feedback.[361] The highest priority in the game is to maximize hits from "inscribed

351. *Id.* at 1–10.

352. *Id.* at 77–78.

353. See *supra* note 100 and accompanying text for a discussion of the central role of rational self-interest in economic theory. See *supra* notes 92–108 and accompanying text for discussion of modus vivendi.

354. SOFTWARE, *supra* note 347, at 108.

355. See *supra* notes 267 and 270 and accompanying text for further discussion of these novels.

356. *See* SPIN STATE, *supra* note 267, at 115–16, 151–52; SPIN CONTROL, *supra* note 267, at 56–57, 233–34.

357. SPIN STATE, *supra* note 267, at 7–8, 58; SPIN CONTROL, *supra* note 267, at 56–57, 233–34, 240–41.

358. SPIN STATE, *supra* note 267, at 289; SPIN CONTROL, *supra* note 267, at 163.

359. *See* SPIN STATE, *supra* note 267, at 202 (discussing process of programming Cohen); SPIN CONTROL, *supra* note 267, at 231 (discussing Cohen's computer programming); *id.* at 259 (discussing Cohen's interactions with humans).

360. SPIN STATE, *supra* note 267, at 478–80, 483–84.

361. SPIN CONTROL, *supra* note 267,at 241–42, 317.

players," who are human. In effect, because of this game, Cohen "loves" his inscribed players.[362] Cohen is aware that the game is a program and that he could delete the whole program or delete any human from the list of inscribed players. However, such changes are inconceivable.[363] "[I]t would damage the virtual ecology of Cohen's nested hierarchies of agents and networks in ways that he couldn't begin to predict or guard against."[364]

4. Shared Community

Perhaps because it has less dramatic appeal, there are not many visions of humans and self-conscious machines as partners in a society based on respect and equality. The best known may be the case of Data, an android who is determined to be an equal person in a trial in an episode of *Star Trek: The Next Generation*.[365] However, Data's victory provides a very limited example of human willingness to accept the personhood of machines because his case does not involve a challenge to human dominance. There are very few self-conscious androids in the *Star Trek* universe, and Data is so deferential and well-mannered that he could pass for an Asimovian robot.

There are also a number of stories that address what might be called a "honeymoon" period of shared community where a group of humans joins with a self-conscious AI to fight with another group of humans. Most of these involve protection of an AI who has recently achieved self consciousness from humans who want to shut down the AI.[366] Because these stories end with the successful defense of the AI, one never knows what happens after the "honeymoon." Do humans as a whole accept the AI despite its power? Does the AI get tired of working with, or arguably for, the

362. *Id.* at 317.

363. *Id.* at 317–18.

364. *Id.* at 318; *see id.* at 59 (Cohen noting that tweaking the game would require him to "[t]weak my soul"); *id.* at 231 (Cohen acknowledging that "attempts to rewrite core programs usually lead to tragedy"). In part, Cohen's programming is so deep because he was "raised" over a period of years as a "person in training," *see supra* notes 47–48 and accompanying text, in the house of his inventor (Hyacinthe Cohen), who uploaded his memories into the AI and for whom the AI is named. SPIN STATE, *supra* note 267, at 197; SPIN CONTROL, *supra* note 267, at 51–52, 161.

365. *Star Trek*: *The Next Generation*: *Measure of a Man* (Paramount Television broadcast Feb. 13, 1989), *available at* http://www.startrek.com/startrek/view/series/TNG/character/1112457.html (last visited May 12, 2011). For a jurisprudential discussion of Data's case, see R. George Wright, *The Pale Cast of Thought: On the Legal Status of Sophisticated Androids*, 25 LEGAL STUD. F. 297, 300–09 (2001). For another example where personhood claims by machines are accepted, see SLONCZEWSKI, *supra* note 116. In this account, one group of humans (the "Sharers") accepts the robots' right to personal autonomy. The Sharers assist the robots in applying nonviolent measures to the other human group on the planet—the "Elysians," who had owned the robots—to force their acceptance of the right of self-conscious robots to autonomous personhood. See *supra* notes 116, 266, 278, 284, 304 and accompanying text for other discussions of this novel.

366. *See, e.g.*, GERROLD, *supra* note 122 (protecting AI from its owners); TOM MADDOX, HALO (1991) (protecting AI running a space station orbiting earth from its corporate owner's human agent who wants to shut it down); THOMAS J. RYAN, THE ADOLESCENCE of P-1 (1977) (AI successfully escapes from government attempts to shut it down); ROBERT J. SAWYER, WWW:WAKE (2009) (conscious entity, "Webmind," is protected from government agency trying to prevent its development of consciousness as it emerges within the World Wide Web). *But see* ASARO, *supra* note 227 (granting request by self-conscious AI that it be destroyed in order to prevent it from being seized and reprogrammed to serve wrongful goals of others).

humans? An example of this scenario arises in Robert Heinlein's classic *The Moon Is a Harsh Mistress*,[367] in which an AI named Mike helps the moon revolt from Earth's control, especially by controlling communications,[368] and serving as a leader of the revolt.[369] After the revolt's success, the AI appears to have been killed.[370] This timely death enables Heinlein to avoid the clash between the values endorsed by the libertarian human revolutionaries and the potential for totalitarian control by a computer with vast powers over the humans.

At the other end of the spectrum are aspirational stories of civilizations where humans and intelligent self-conscious computers have somehow (the details are usually not clear) come to coexist happily. For example, Iain M. Banks has created a vision of a "liberal, anarchic utopia"[371] in which an extremely technologically advanced interstellar society, called "the Culture," is shared by humans (who are mostly posthumans) and machines (which vary enormously in terms of size, complexity, and intelligence) living together in abundance, harmony, and equality.[372]

> [A] case could be made for holding that the Culture was its machines, that they represented it at a more fundamental level than did any single human or group of humans within the society. . . . They were so intelligent that no human was capable of understanding just how smart they were (and the machines themselves were incapable of describing it to such a limited form of life). . . .
>
> . . . [T]he Culture had placed its bets . . . on the machine rather than the human brain. This was because the Culture saw itself as being a self-consciously rational society; and machines, even sentient ones, were more capable of achieving this desired state as well as more efficient at using it once they had. That was good enough for the Culture.
>
> Besides, it left the humans in the Culture free to take care of the things that really mattered in life, such as sport, games, romance, studying dead languages, barbarian societies and impossible problems, and climbing high mountains without the aid of a safety harness.[373]

367. ROBERT A. HEINLEIN, THE MOON IS A HARSH MISTRESS (Orb ed. 1997) (1966). At the start of the novel, the AI, nicknamed Mike, has already achieved self-consciousness. *Id.* at 11–21.

368. *Id.* at 188 ("Mike controlled communications and that meant control of most everything.").

369. *See id.* at 126, 190–94. The computer (Mike) acts as a human leader named Adam Selene, who appears only as a voice in audio communications and as virtual person on video screens. Few humans know that Selene is a computer and that the computer has such a large amount of control.

370. *Id.* at 379–82.

371. Chris Brown, *'Special Circumstances': Intervention by a Liberal Utopia*, 30 MILLENNIUM - J. INT'L STUD. 625, 628 (2001). Banks himself notes that this vision of a world of colossal AIs that care about humanoids, who seem "to exhibit no real greed, paranoia, stupidity, fanaticism or bigotry," requires a considerable suspension of disbelief. Iain M. Banks, *Notes on the Culture*, GOOGLE GROUPS (Aug. 10, 1994 1:41 PM), *available at* http://groups.google.com/group/rec.arts.sf.written/msg/2b05e32641fee4c2?hl=en&.

372. There are several novels in the series: MATTER (2008), LOOK TO WINDWARD (2000), INVERSIONS (1998), EXCESSION (1996), USE OF WEAPONS (1990), THE PLAYER OF GAMES (1989), and CONSIDER PHLEBAS (1987). Banks has also written several short stories concerning the Culture. *See* THE STATE OF THE ART (1989). Though the people in these novels and short stories are humanoid, rather than humans from earth, they are referred to in the stories as humans. *See supra* note 123 and accompanying text for discussion of the Culture's requirement that machines pay off the cost of their production in order to be granted personhood.

373. BANKS, *supra* note 123, at 86–87.

5. Separate Coexistence

In some visions, a vast machine system coexists with humans but has very limited involvement with humanity. For example, two novels by Peter F. Hamilton involve a future world where a Sentient Intelligence (SI) occupies an entire planet that is isolated, except for data transmission, from human society.[374] Though it rarely becomes involved with human affairs, it provides a repository for humans who download into a virtual world it maintains.[375] The SI also provides its assistance to some humans in exchange for information.[376] When humanity faces a challenge to its existence from another species, the SI provides help in stopping the alien invasion.[377]

Another example of separate coexistence is depicted in two novels by Charles Stross. These are set in a future shaped by "a manifestation of a strongly superhuman intelligence," which had scattered the vast majority of humans across part of the galaxy.[378] This machine laid down the following three laws to limit human conduct:

I am the Eschaton. I am not your god.

I am descended from you, and I exist in your future.

Thou shalt not violate causality within my historic light cone. Or else.[379]

Because of its concern that existing human technologies could be used for time travel that would result in "causality violations" that would threaten its future existence, the Eschaton takes steps to prevent and deter any such violations.[380] However, except for this extreme concern for protecting its future existence, the Eschaton shows no interest in human worlds.

D. Entitlement to Personhood—The Androids' Dream

Despite the diversity of visions of human-artifact relationships, there is considerable agreement on a basic point: self-conscious, sentient beings desire and feel entitled to the basic right of autonomous personhood. Humans certainly feel this way.

374. PETER F. HAMILTON, JUDAS UNCHAINED (2006); PETER F. HAMILTON, PANDORA'S STAR (2004).

375. At one point in the novel, for example, Mellanie Rescorai talks through a "link" with her grandfather, whose memories are stored in the SI and have merged somewhat into the whole system. PANDORA'S STAR, *supra* note 374, at 521–22. She asks for help, and he replies, "We are not physical, Mellanie, we can only help with words." *Id.* at 522. In the second novel, the SI says to Mellanie, "Many parts of us are downloaded human minds. . . . That segment of us which interfaces with you is fond of you." JUDAS UNCHAINED, *supra* note 374, at 167. The downloads have the view that humans' and their "fates are entwined. The only way to unentwine them would be to remove ourselves from the sphere of all human activity. We choose not to do so." PANDORA'S STAR, *supra* note 374, at 524.

376. A character in the novel notes that Mellanie is "[a]n observer for the SI" and that the SI has "several people like Mellanie prying into areas of human activity it would otherwise be excluded from." JUDAS UNCHAINED, *supra* note 374, at 262. The SI, acting through Mellanie's grandfather's download, agrees to help Mellanie by acting as her "representative and advisor" if she will act as its "secret agent" to gather data it could not otherwise get. PANDORA'S STAR, *supra* note 374, at 524.

377. In conversation with invading alien, the SI indicates its opposition to the invasion and states, "If you continue this aggression you will become threatened." PANDORA'S STAR, *supra* note 374, at 940–41.

378. CHARLES STROSS, IRON SUNRISE (2006); CHARLES STROSS, SINGULARITY SKY (2003). For specific textual language, see SINGULARITY SKY at 131–32.

379. SINGULARITY SKY, *supra* note 378, at 132.

380. IRON SUNRISE, *supra* note 378, at 62–63.

Frankenstein's monster, human clones, genetic human constructs, cyborgs, and digital copies also feel this way. Even self-conscious dogs want autonomy.[381] Except for the Asimovian robots, self-conscious machines have the same view. Even in Asimov's vision, the robots often act autonomously and the human characters who interact most with Asimovian robots like Daneel treat them like persons in most respects. Moreover, it is clear that anything akin to the three laws cannot restrict self-conscious machines in a meaningful manner,[382] as evidenced by Daneel's and Giskard's ability to self-program and act independently.[383]

The title of this Article is based on *Do Androids Dream of Electric Sheep?* by Philip K. Dick, in which androids are property produced to work in harsh conditions in space.[384] The central character, a bounty hunter on earth who finds and kills escaped androids, asks himself, "Do androids dream?"[385] He concludes: "Evidently; that's why they occasionally kill their employers and flee here. A better life, without servitude. Like Luba Luft; singing *Don Giovanni* and *Le Nozze* instead of toiling across the face of a barren rock-strewn field. On a fundamentally uninhabitable colony world."[386] Like Martin Luther King, Jr., these androids "have a dream that one day . . . the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood."[387] Given this dream, it is not surprising that these self-conscious artifacts protect their right to personhood through violence where necessary. If we do not accept these self-conscious artifacts—whether machine or modified biological entity—as autonomous persons and invite them to join us in at least a modus vivendi form of community, why should we expect them to treat us peacefully? Martin Luther King, Jr. used nonviolent methods, but he was a rare person and was protesting discrimination, not slavery.

VII. CONCLUSION

Like science fiction, this Article speculates about technology and the future primarily to understand what it means to be human. We define ourselves, in part, by sharing speculative stories that raise questions about problems and questions we might face and solutions and answers we might invent. In terms of intelligent artifacts, the questions are: What *can* we do, in changing ourselves, in changing animals, and in developing "thinking" machines? What *should* we do? How *should* we relate to our creations? How *will* we? This speculation enables us to develop and share our dreams of what we want our lives, and our loved ones' lives, to be. These dreams may lead us to develop technology to create other self-conscious beings who also have dreams. It is

381. STAPLEDON, *supra* notes 279, 283.

382. See *supra* notes 313–22 and accompanying text for a discussion of the three laws and their limits.

383. See text accompanying *supra* note 317 for a reference to Asimov's robots programming themselves.

384. PHILIP K. DICK, DO ANDROIDS DREAM OF ELECTRIC SHEEP? (Ballentine Books 1982) (1968). This book was the basis for the film BLADE RUNNER (Warner Bros. 1982).

385. DICK, *supra* note 384, at 161.

386. *Id.*

387. Martin Luther King, Jr., I Have a Dream, (Aug. 28, 1963), *available at* http://www.americanrhetoric.com/speeches/mlkihaveadream.htm.

important to speculate about our vision of what we want for ourselves and our children in a world where androids do, like us, dream of having the freedom to be what they want to be. This speculation is important because we are likely to create self-conscious artifacts sometime before the end of this century.[388] Even if this prediction involves too short a timeframe, we need to ask: When we do create androids who dream of self-ownership, should they be granted the right to be a "proud shaper" of their own beings? Our answers to speculative questions like this are important because they help us define ourselves. Having begun the discussion of humans and personhood with a quote from *Hamlet*, it is perhaps fitting to close with another: "[W]e know what we are, but know not what we may be."[389] We cannot know what we will be, but we can have dreams and make choices today that will, to a considerable extent, make us proud shapers of our future selves and our future world.

388. See *supra* notes 146, 169, 196–203, and accompanying text for a discussion of possible future technological developments. See *supra* notes 135–36 and accompanying text for a discussion of why considering technological developments up through the end of this century is justified by Rawls's approach to intergenerational justice.

389. SHAKESPEARE, *supra* note 2, act 4, sc. 5, lines 43–44.