Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Amy Michelle DeBaets	4/10/12
Student Name	Date

The Robot as Person:
Robotic Futurism and A Theology of Human Ethical Responsibility Among Humanoid Machines

By
Amy Michelle DeBaets
Doctor of Philosophy
Graduate Division of Religion
Ethics and Society

Elizabeth Bounds
Advisor

Pamela Hall
Advisor

Paul Wolpe
Committee Member

Ian McFarland
Committee Member

Accepted:

Lisa Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date:

The Robot as Person:
Robotic Futurism and A Theology of Human Ethical Responsibility Among Humanoid Machines

By
Amy Michelle DeBaets
Th.M., Princeton Theological Seminary, 2007
M.Div., Princeton Theological Seminary, 2006
M.A., Trinity International University, 2000
B.S., Truman State University, 1997
B.A., Truman State University, 1997

Advisors: Elizabeth Bounds, Ph.D. and Pamela Hall, Ph.D.

An abstract of a dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate Division of Religion, Ethics and Society, 2012.

Abstract

The Robot as Person:
Robotic Futurism and A Theology of Human Ethical Responsibility Among Humanoid Machines

By
Amy Michelle DeBaets

A world in which humans and robots coexist is one with tremendous possibilities for good and ill. Futurist thought in robotics has contributed both positively and negatively to the development of humanoid robots to this point, offering ideas and values about what it means to be human and what it could be for a robot to be a moral person. Some of the more popular forms of robotic futurism have tended to overemphasize intellection and a disembodied mind as the ultimate form of existence, while the more constructive forms have looked at human emotional and social interactions and patterned their robots after them. Robots that are embodied, sociable, and situated in their environment and history are ones that begin to mirror humanity and the beings that we consider to be morally valuable in themselves. But robotics and related psychology do not offer a complete picture into the possibilities for robotic personhood in interaction with human beings. It is here that theology can provide a useful history of reflection and understanding of personhood beyond the human that can begin to develop creative possibilities for the future direction of robotic personhood as well. Fully humanoid robots, then, could embody the qualities of freedom and constraint, goodness and fallenness, finitude and transcendence, and embodied spirituality that characterize human personal life. These qualities can be considered in the development of robustly humanoid robots in a number of different application areas and the ethical effects of those developments can be better understood using these criteria. Humanoid robots can perform jobs that humans cannot or would not do, they can change the ethical calculus of war, and they may even be able to provide genuine companionship and friendship to human beings, but they need to be designed in such a way as to facilitate human flourishing first, so that robotic flourishing can follow.

The Robot as Person:
Robotic Futurism and A Theology of Human Ethical Responsibility Among Humanoid Machines

By
Amy Michelle DeBaets
Th.M., Princeton Theological Seminary, 2007
M.Div., Princeton Theological Seminary, 2006
M.A., Trinity International University, 2000
B.S., Truman State University, 1997
B.A., Truman State University, 1997

Advisors: Elizabeth Bounds, Ph.D. and Pamela Hall, Ph.D.

A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate Division of Religion, Ethics and Society, 2012.

Table of Contents

Acronym List

AI - Artificial Intelligence

AAAI - Association for the Advancement of Artificial Intelligence

ASIMO - Advanced Step in Innovative Mobility (Honda-developed robot)

DARPA - Defense Advanced Research Projects Agency (US)

DHS - Department of Homeland Security (US)

DOA - Department of Agriculture (US)

DOD - Department of Defense (US)

DURIP - Defense University Research Instrumentation Program (US)

IEEE - Institute of Electrical and Electronics Engineers

LoAR - Law of Accelerating Returns

NASA - National Aeronautics and Space Administration (US)

NIH - National Institutes of Health (US)

NIST - National Institute of Standards and Technology (US)

NRI - National Robotics Initiative (US)

NSF - National Science Foundation (US)

RLE - Radical life extension

ROE - Rules of Engagement

RUR - *Rossum's Universal Robots* (play)

WHO - World Health Organization (UN)

Introduction

This is a story about the future.

The human imagination has long been fascinated with creating and using tools by which we adapt our environments to our wants and needs as well as the idea of creating a new version of ourselves, not through ordinary reproduction, but rather fashioned by our own hands, using our minds and manipulating nonliving materials. Virtually all the contemporary literature on the history of robotics notes this development: from the ancient myths of Pygmalion and the golem to more modern visions of Frankenstein's monster and the worlds of cyberpunk, human-like beings made by human hands have been the stuff of dreams and nightmares. We are both attracted to and repelled by the possibilities of these creations. We envision that they might make human life easier and that they might rebel and overthrow us. They might bring about either utopia or dystopia.

These visions of humanoid creations have heretofore been limited to the realm of myth. Until very recently, science fiction has been more fiction than science. But science is beginning to catch up to the fiction, and the fields of robotics and artificial intelligence are rapidly working toward the development of lifelike, intelligent, adaptable machines that can explore their worlds and interact with humans in new and interesting ways. We are still a long way off from creating fully humanoid machines, but robots now walk, talk, clean floors, deactivate landmines, and even perform weddings.

Despite the growing presence of robotics in contemporary life and popular culture, theologians have had little to contribute to the discussion, whether reflecting on ethical issues in robotics or their religious implications. Only a few isolated works exist, and

these primarily examine the implicit soteriologies and other religious visions in popular robotics literature, the contribution that robotics and AI could make in the science / religion dialogues, or a creation / creativity-centered theological anthropology of current robotics work.[1] There is, on the other hand, a great deal of reflection on robotics outside the theological world, which I believe could be strengthened by substantive theological consideration of robots within theological anthropology and ethics. This theological consideration takes seriously the embodied nature of human life, its emotional and spiritual content, the relationships and circumstances in which we are embedded, and the polarities of action, such as finitude and transcendence, freedom and constraint, which define our everyday existence.

The primary question that I seek to explore in this project is what makes a being a "who" rather than a "what," a person rather than a thing, a moral subject. I do this through a consideration of contemporary movements to develop humanoid robots, looking critically at the anthropologies underlying such attempts, exploring the ethical and practical implications of those underlying understandings of human (and robotic) personhood, and using robotic futurism and Christian theology to begin to develop a constructive narrative to further develop an understanding of the possibilities for humanoid robots. I am seeking to tell a story about the future of human beings' development of and interaction with robots. Robots serve as an interesting and fruitful point at which to explore ongoing questions of human nature – what we think a person is and why influences how we develop and utilize robots.

[1] These will be explored more fully in Chapter 1.

The project will begin by introducing several key questions about the nature and trajectories of current developments in robotics. What makes a being a person, and why? How does how we think about what makes someone a person influence how we develop humanoid robots and related technologies? What are the ethical implications of our anthropological understandings for the interactions between humans and our robots? And how might theology contribute constructively to the conversation?

As more advanced humanoid robots come to be developed, the context of human moral responsibility changes somewhat to include ethical consideration for the robots themselves as potential moral beings to whom and for whom we have responsibility. How and why humanoid machines might be thought of and treated as persons / moral beings / "human," etc. will begin first within the theological anthropology developed and then considered fully within the contexts in which they could be developed and exist. This will include an analysis of how robots might come to be moral agents themselves within their own contexts.

Robotic futurism and artificial intelligence have long been plagued by unrealistic expectations of what lies "just around the corner." Futurists have predicted the development of intelligent humanoid robots, yet the science and technology of robotics has lagged well behind the predictions, even as other areas of technology have seen rapid advancement beyond anyone's wildest expectations. The reasons for this, I believe, lie in the trajectories that robotics and AI development have taken in working to solve the problem of what constitutes "intelligence" in AI.

For a long time, in the early days of AI, true "intelligence" was understood to be

what roboticists found challenging – primarily symbolic manipulation problems like chess. Things that most two year-olds can do, like walking, talking, and relating to other people, were not considered "intelligent," yet these have been some of the most intractable problems in robotics. Some current work in AI and robotics has developed a different understanding of what constitutes robotic intelligence – embodied, sociable, situated beings that can learn to interact successfully with the world around them, including interacting with human beings.[2] I believe that this constitutes a good beginning for developing an understanding of personhood - an embodied, contexual, relational existence in which a being learns and grows in interaction with others and the environment.

I do not presume to be a futurist, though I will explore what I believe are some of the likely trajectories of emerging work in robotics. This will involve an analysis of what has taken place in a variety of areas of robotics and some of the technical roadblocks within applications development. The analysis will be situated in the realistic problems and circumstances of human life, and it will be based in what is currently being developed and funded. In this, I hope to avoid some of the problems faced historically by futurism. For instance, in 2000, Hans Moravec predicted that by 2010, robots would clean people's houses, including applications like vacuuming and dusting. We now have robots that vacuum (my Roomba cleans my floors automatically every few days), but the technical challenges of moving from vacuuming (which operates essentially in two-dimensional space and involves no manipulation of objects) to dusting (which involves moving in three

[2] Rodney Brooks. *Flesh and Machines: How Robots Will Change Us* (New York: Vintage, 2003).

dimensions, as well as significant manipulation of objects within that space) went utterly unexplored by Moravec. He simply assumed that "cleaning" was, more or less, one relatively simple set of applications. I hope to be realistic in understanding the technical bottlenecks that exist in developing certain applications.

As mentioned, one of the key factors in considering future trends in robotics and AI development is research and development funding. In Japan, much of the current funding stream for robotics is geared toward personal service applications for an aging society. In contrast, in the US, more of the funding for robotics is directed toward military and defense applications.[3] These very different sources of investment in robotics and intended projects lead to very different sets of applications being developed, and this should be taken into account when considering long-term trends in the field. Robotics development is expensive and takes time; major breakthroughs are unlikely to come from outside of the mainstream funding sources.

The questions to be addressed in contemporary robotics are, for the most part, in line with a trajectory of technologies that have been under development and causing changes in human culture for some time now. Roombas are not qualitatively different from other forms of automation in manufacturing, service, and information that have become increasingly important since Guttenberg invented the printing press. The development of advanced robotics continues that trend, and these questions have become

[3] For instance, Ray Kurzweil serves as a primary advisor to the US Army on its development of humanoid robots to serve as "tactical autonomous combatants" – robots that decide autonomously when and whether to shoot. (Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology* (New York: Viking, 2005), 333).

ever more important as machines come to be relied upon for critical decisions, but history can and should be instructive. Many of the questions involved in theological anthropology and ethical responsibility are not new, even if their applications are somewhat novel.

Historical trends should be taken into account when looking at ethical questions in robotics and AI, but potential qualitative shifts in the field must be considered as well. There may come a time when an intelligent, learning, thinking, feeling robot, who can interact with human beings in ways very much like other humans, exists. What might constitute such a qualitative shift, and even the language with which to speak of that shift, will be addressed. I hesitate to speak of the qualitative shift as the development of "intelligent" machines, out of concern that such language might answer the question before it is asked. Likewise, to speak of "autonomous," "emotive," or even "situated and embodied" robots is to determine in advance what the critical factors are. The best language I have at this point is to speak of "humanoid" robots, which I think leaves the key anthropological questions sufficiently open.

I became interested in studying what it means to be a "person" in the course of my early work on bioethics, particularly in its application in emerging technologies. What is it, morally, theologically, politically, and socially to be a human being? How does the "human" relate to the "person"? And why do we differentiate between the two? As often happens in life, my career took a detour for several years into the field of information technology, during which time I found that the convergence of IT with biology and cognitive science would be having a tremendous impact in the coming years on the shape and form of human life. I entered the field as internet access was just beginning to be

widespread, and by the time I left seven years later, the internet was an absolutely necessary facet of everyday life - broadband was increasingly available, and everything from phone calls to banking to libraries was becoming mediated through digital technologies. As I studied the great changes happening in the understanding of human personal identity with advances in genetics and neuroscience, I found that it was often difficult to have a constructive conversation around them because of the politicization of the science and the questions involved. Was human nature something that was absolutely inviolate, or could it and should it be modified through genetic technologies? This project came about as a means through which to explore questions of personhood in a way that avoided the most problematic politicizations, in a field that is asking how to create something like a human being, rather than changing the humans that currently exist. So, while robots are the primary topic of this dissertation, they are, in a real sense, only an example of a broader set of questions about the moral, theological, and political meaning of "personhood," how we understand it, how we reflect it in our technological development, and how we treat one another in the process.

I would like to give special thanks to the members of my committee for their continual support and encouragement throughout this project - Pam Hall, Liz Bounds, Paul Wolpe, and Ian McFarland. Thanks also to Jessie Smith, Rez Pullen, and Nikki Karalekas, who all read the proposal and some early drafts. Special thanks to Cyd Cipolla and Kristina Gupta who read and commented on many drafts of the various chapters here, and without whom this project would not be complete. Thanks to Jennifer Thweatt-Bates for being a great colleague and friend in the field. And much love and thanks to Brian for

always tirelessly supporting and caring for me and for proofreading and copyediting this dissertation.

Chapter 1: Key Questions About Humanoid Robotics and Ethics

Of Humans and Our Machines

Machines that mimic human functions have held a place in the human imagination across a variety of times and cultures, but the ability to create such machines is relatively recent. Popularized in science fiction throughout the 20th century, the term "robot" was coined by Czech playwright Karel Capek in 1920 in his *RUR (Rossum's Universal Robots)*, a play about human-created machine life forms that gain souls (creatively defined as the ability to hate) and revolt against their human masters and the meaningless drudgery of their existences.[4] The term "robot" itself arises from the Czech *robota,* meaning drudgery or forced labor. While Capek's play was popular both in Europe and the United States when it appeared, it was not until the 1950s that robots fully entered the popular lexicon and imagination with the debut of Isaac Asimov's *Robot* series of books. Since that time, robots have flourished within the world of science fiction and have become a target for development within information technology and artificial intelligence circles.

In their most basic form, robots are machines that are designed and programmed to do automated physical work of some kind, but this remains a very open definition that can encompass more than just the machines that we think of as robots. More specifically, robots manipulate physical objects within their environment and have at least some programmed mobility of their own. A robotic arm is one that could grasp an object and

[4] Karel Capek, *RUR (Rossum's Universal Robots)* (New York: Penguin Classics, 2004).

change it in some way, which is differentiated from a machine that does not itself move but still manipulates objects (such as a traditional manufacturing machine), or a simple computer (which manipulates data but not physical objects). The robot, to perform its tasks, must use computing technology in order to move within its environment. Robots are then generally used for some kind of labor, whether that labor is building cars, cleaning floors, defusing bombs, or playing the violin.

But the ambition of roboticists has never simply been the automation of various kinds of labor. From the founding of the field within science fiction in the work of Capek and Asimov, a key idea behind the development of machines that can interact in complex ways with their environments has been the development of machines that duplicate the full range of human characteristics - to make machines that are, in a real sense, people. Capek's robots had been happy with their lives of drudgery until their souls were awakened, and they yearned for the freedom that their human masters experienced. Asimov's robots likewise had the power to reflect on their own existence, work, relationships, and responsibilities. They were programmed with a basic ethical sense to protect humans and themselves, and in so doing, highlighted the flaws in human ethical sensibilities.[5]

One of the key questions that I seek to explore is what it means for a robot to be a person - to be humanoid. This requires a deep inquiry into what characteristics qualify such a machine as a "who" rather than a "what." But it is not simply a matter of defining the "right" set of characteristics and applying them accordingly. Robotic futurists have

[5] Isaac Asimov, "The Evitable Conflict," *I, Robot* (New York: Random House, 1991).

struggled with, and sometimes simply assumed, what those characteristics were, and those struggles and assumptions reflect a particular set of cultural, political, and philosophical dynamics that have real ethical consequences, both for the machines that are developed and for a broader understanding of what it means to be a human person as well. Because of the importance of human interactions with our machines, the choices made in the development of those machines have long-term consequences for the future of humanity. The values and assumptions reflected within those choices should not solely be those of a small group of elite technofuturists, but should rather be part of a broader public conversation about what it means to be human, what the future of human-machine interactions should be, and what choices will actually contribute to the flourishing of all human beings as well as the machines that we create to be like us. Our robotic anthropologies[6] both reflect and have the power to adapt what (and whom) we value within human beings as well. Where our robotic anthropologies are reductionistic and exclusionary they can influence how we treat the humans who interact with and are replaced by those robots, and where they are morally creative and expansive they may be able to open up new futures that create a better world for human and robot alike.

I have described these robots as "humanoid" up to this point, as it is the human that serves as the standard and point of comparison for the development of robots as persons. I intend to use the term in such a way as to neither overdefine the object of development from the start, nor to assume any particular anthropology, but to analyze

[6] Throughout my work I will use the term "anthropology" in its classic theological and philosophical sense to denote the study of what it means to be a human being and not in its disciplinary sense within anthropology for the study of human cultures.

working anthropologies and to build toward a working definition as I go. I am choosing not to use terms like "artificial intelligence" or "android" to describe such robots, though they are traditional terms that have been used, but each builds in particular values and assumptions that I hope to avoid in exploring what it means for a machine to be a person. The term artificial intelligence (AI) reflects the value that what "really counts" in a humanoid robot is the fact of intelligence, which, while important, obscures issues like embodiment, emotion, and ethics. Android, on the other hand, connotes a specifically gendered form of robot (aner-, Greek for male or masculine), and thus will only be used when the genderedness of the robot matters.[7] I will likewise avoid the traditional (pre-*RUR*) term *automaton*, as it implies a lack of freedom and openness for the robot that we will likely find to be important in understanding machines as persons. One term that I believe can be used interchangeably with humanoid in reference to robots as persons is *anthropic* (person-like), which can be most useful in comparing robots to non-human persons, but I will not use this as a primary term because it can be mistakenly connoted with the anthropic principle in cosmology.[8]

While I do not want to overdefine the concept of the humanoid from the start, there are several features that may be important in thinking about the full range of

[7] Some thinkers in the field have used the term "android" to refer to humanoid robots; I will not use this because of its etymological connotation of necessary maleness. Android, and its corollary, gynoid, will only be used in reference to robots that serve particular sexual functions, or where the gender of the robot matters.

[8] The anthropic principle is an idea within cosmology that the universe is set up in such a way that either allows for, or in its strong version, compels, the development of intelligent life that can reflect on its existence within that universe (and thus, develop something like an anthropic principle). The point is not to debate the principle, but to avoid confusion with it.

possibilities in what may make a given robot a "who" and not a "what." A basic feature that is not intrinsic to humanoid versions of robots but to robots in general in relation to other complex computational devices is that of mobility and engagement with the physical universe. While it can (and has) been argued that future persons could exist in strictly virtual environments, physical mobility and manipulation of the world are key to defining a machine as a robot and will thus be part of the definition of a humanoid robot.[9]

One feature that roboticists have assumed would be a feature of any humanoid robot is that of intelligence and independent decision-making and action. While this is certainly important, I seek to avoid the mistake of also overdefining intelligence within this context. A key problem within the history of robotic futurism that will be explored in the next chapter is that "intelligence" has tended to be defined within the robotics community as those activities that technophilic programmers have found difficult or interesting. There must be room within our definition of the humanoid for a broader understanding of intelligence and its place within the overall concept of the person, whether robotic or human, such that "intelligence" is not reduced to the ability to meet a narrow set of intellectual challenges and that "person" is not reduced simply to intelligence.

An aspect of humanoid robots that is related to, but not quite the same as, intelligence is the ability to learn and change. Learning will certainly include adding to intelligence, but it will also include adaptation to the environment, emotional responses to other people, the ability to choose what one learns, and improvement of ethical decisions

[9] This is a key feature of both Kurzweil and Moravec's works, among others.

and actions, among others. I want to keep the idea of learning, like that of intelligence, as open as possible in exploring what it means to be a person and to begin to discover the shape of the concept within the broader context of understanding persons.

Some may believe that robotic personhood is primarily a matter of consciousness, but I will not attend to questions of consciousness directly. The problem of consciousness in both philosophy and neuroscience is that consciousness and self-awareness are subjective states that are not empirically testable or falsifiable. That is, I can never know whether a being other than myself is actually conscious, or whether the being merely appears to be conscious. I will instead attend to some of the empirically identifiable correlates of consciousness, such as sociality and adaptive learning. Roboticists like Cynthia Breazeal have highlighted the importance of interactivity and sociability within robots as a crucial aspect of humans relating to robots as persons instead of things.[10] The relational, emotional, and empathetic aspects of human life may also be considered important in understanding machines as persons, though this has not been a common feature that has been valued by earlier generations of roboticists and futurists. Breazeal's work with sociability in robots has highlighted the importance of the ways that humans respond to the robots that we design, that part of the "personhood" of a robot may be in whether we treat it as such, or whether we treat robots as things. Her work has shown that the features that lead to humans treating robots "personally" are not primarily intelligence and learning but facial responsiveness and physical interactions, regardless of whether the robot is otherwise "intelligent."

[10] Cynthia Breazeal, *Designing Sociable Robots* (Cambridge, MA: MIT Press, 2002).

Yet another aspect of robotic personhood that will be explored is that of the freedom that currently differentiates humans from our machines, and humanoid robots from their automaton predecessors. The idea of an open future in which one can make choices, learn, act, and respond, in which one has responsibility for oneself and one's relationships with others, is important to explore in understanding robots as persons and is reflective of our understanding of our human lives as "personal." This reopens classic discussions about free will, determinism, and the relationship of humans to the larger universe, for this very reason it must be explored in the robotic context as well. I will explore what it could mean for a robot to be both the object and subject of moral action in the context of its situated embodiment and emotional relationships with others.

What we develop in our humanoid robots and how we make use of those robots raises a variety of interesting and important ethical questions as we continue to interact with our machines in ever more personal ways. We ought to ask why we develop what we do in our machines, what aspects of robots make them "humanoid" and why, and how we should treat robots that are considered to be "persons." This question of what makes a being "personal," a "who" rather than simply a "what," a subject and not solely an object of action, will be the primary question explored throughout this work.

Why the Contributions of Theology?

I believe that there is a critical role to be played by theology in the shaping of future humanoid robotics development for several key reasons. First, theology provides a long history of critical reflection, contention, and refinement of understanding of what it

means to be a person, ethically and theologically. Christian theologians have argued for centuries about the respective natures of various kinds of being and what the ethical treatment of those various kinds of beings should be. Unlike most philosophical considerations of personhood, theological reflection has never thought of humans as the only (or even primary) beings that are persons. As I will show, in Christian theological thought, humans are persons by image, and there is space available in creation for both human and non-human persons.

The second is that robots are, to a great extent, a reflection of ourselves and what we value in ourselves. We focus on building that which we value, what we find challenging, what we need help with, and what we yearn to be. We begin by attempting to create new versions of ourselves through building what we value in ourselves; the history of humanoid robotics development mirrors some popular Christian anthropologies that identified the image of God with human rationality. Just as classical Greek philosophy deemed humanity the "rational animal," so also early robotics development emphasized the rational, symbol-manipulating, logic problem-solving aspect of human intelligence. While an emphasis on machine intelligence and problem solving remains, many of the emerging trends in humanoid robotics development focus instead on building relationships between humans and our machines. Behind the movement to develop sociable robots is an understanding of humans as fundamentally sociable and emotional animals, as well as a desire to understand the human psychology behind the development of our relationships. Likewise, the drive to begin to build ethics into the design of machines reflects understandings of the place of morality in human existence. It may be that theological

reflection may offer a basis for a next phase of robotics development, in which robots are designed in order to be transformed into closer reflections of goodness and relationship to other creatures (and God).

Third, I believe that theology is important because it brings a critical realism in both anthropology and political life. Christian thought has long considered what it is to create something, what the possible or necessary trajectories of human history are, and the meanings of apocalypticism, embodiment, death, and hope. Theology has worked to avoid political and social naivete while retaining a sense of hopefulness and a vision for the future of humanity and all creation. Christian reflection provides a set of visions for what real moral progress looks like without assuming it will happen as part of an inevitable trajectory of history. It acknowledges human sinfulness and tendencies toward greed, self-justification, pride, and lust for power while also claiming human goodness and potential for faith, hope, and love. At its best, theology provides a critical realism for recognizing how much we do not know, the limits of our knowledge, and the frailty of human life, knowledge, and goodness. Most importantly, Christian theology provides a lens through which to conceive of grace - the creation of goodness where it does not exist.

Fourth, Christian theology can add insights to the picture of human (and robotic) personhood developed in the best futurism - the vision of persons as embodied, sociable, situated beings with emergent minds that can learn, emote, and plan. We can recognize the real tensions that also characterize human life as embodied, spiritual creatures - finitude and transcendence, goodness and fallenness, freedom and constraint - and know that these will be true of anything that we make as much as it is true of us as persons. Our creations

are not likely to be any morally or spiritually better than we are; as human products they could be little else. But as we reflect on these facets of human life we can be more conscious (and conscientious) about how we design and implement them within our machines.

Finally, I believe that theology can provide a robust reading of a pluralistic world in which humans and non-humans, including humanoid robots, can live and thrive together. Christian claims about the nature of persons, politics, and the trajectories of history are made from within a particular, concrete faith tradition and framework of understanding but can be applied and understood in creative ways in contexts beyond itself. The insights of theology can be useful in a society that does not share that theology. Theology can claim an ethical value for difference within its anthropology that contributes to the flourishing of all. I do not believe that one need accept fully a theological account of human life in order to accept that theology can add something substantive to the picture of what it means to be a person. While it is certainly a stretch to say that any idea of humans being made in the image of God would be recognizable outside an explicitly theological context, the understanding of the person as embodied and spiritual, finite and transcendent, good and fallen, free and constrained, is one that has broader applicability. A picture of the human that includes all of these conditions is one even a non-theist could appreciate. What may be challenging, but I believe has real relevance, is the ethical vision of grace that theology brings to the table in considering the future of non-human personhood, specifically the development of a robot as a moral being. To create goodness where none exists as the ethical imperative driving toward the

flourishing of all creation - within this can be beauty, restoration, relationships, and virtue, and I believe that it is a vision toward which we should strive in that development.

Theological Reflections on Humanoid Robotics

Although theology has the potential to make significant contributions to the development of robotic anthropology, there has been relatively little theological reflection on anthropology and robotics, and much of that reflection has centered on a creation-based epistemology and anthropology. The first group of these reflections includes work on robotic theological anthropology, particularly in the work of Noreen Herzfeld, Russell Bjork, and Antje Jackelen.[11] These works reflect on robotic anthropology in ways that I believe are helpful but ultimately insufficient for developing a robust and constructive robotic anthropology. The second group involves work that addresses questions of anthropology but primarily seeks to integrate robotics into current "science and religion" conversations, particularly Ian Barbour and Anne Foerst.[12] The third group considers the theological impacts of robotics and robotic futurism on popular culture but does not

[11] Russell C. Bjork, "Artificial Intelligence and the Soul," *Perspectives on Science and Christian Faith* 60, no. 2 (June 2008): 95-102. Noreen Herzfeld, "Creating in Our Own Image: Artificial Intelligence and the Image of God," *Zygon* 37, no. 2 (June 2002): 303-316. Noreen Herzfeld. *In Our Image: Artificial Intelligence and the Human Spirit*. (Minneapolis: Fortress Press, 2002). Noreen Herzfeld. *Technology and Religion: Remaining Human in a Co-created World*. (Conshohocken, PA: Templeton Press, 2009). Antje Jackelen, "The Image of God as Techno Sapiens," *Zygon* 37, no. 2 (June 2002): 289-302.

[12] Ian Barbour, "Neuroscience, Artificial Intelligence, and Human Nature: Theological and Philosophical Reflections," *Zygon* 34, no. 3 (September 1999):361-398. Anne Foerst, "Artificial Intelligence: Walking the Boundary," *Zygon* 31, no. 4 (December 1996): 681-693. Anne Foerst. *God in the Machine: What Robots Teach Us About Humanity and God.* (New York: Dutton, 2004).

directly address questions of anthropology, particularly in its apocalyptic implications, in the work of Robert Geraci, Michael DeLashmutt, Matt Rossano, and Laurence Tamatea.[13] Outside of particular work on robotics, there have been some interesting developments in theological anthropology in relation to other technologies, particularly technologies that modify or adapt existing humans through technological means. The works of Philip Hefner and Brent Waters offer some constructive possibilities that can be applied to robotics because of the ways in which they integrate insights that are raised in some contemporary robotic futurism with different strains of Christian theology.

Noreen Herzfeld's work on robotic anthropology concerns whether humanoid robots would be considered to be made in the image of God. She examines three interpretations of the meaning of the *imago Dei* in human beings as "reason, regency, or relationship" and their correlates in the development of artificial intelligence as the image of the human (*imago hominis*).[14] Her goal is to

> address some questions that our desire to create an intelligent computer raises about out own nature, namely, why we might be so interested in creating an artificial intelligence and what the approach we take to doing so reveals about our nature as human beings…What, precisely, are the qualities or capabilities that we

[13] Michael W. DeLashmutt, "A Better Life Through Information Technology? The Techno-Theological Eschatology of Posthuman Speculative Science," *Zygon* 41, no. 2 (June 2006): 267-287. Robert M. Geraci, "Apocalyptic AI: Religion and the Promise of Artificial Intelligence," *Journal of the American Academy of Religion* 76, no. 1 (March 2008): 138-166. Robert M. Geraci. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality.* (New York: Oxford University Press, 2010). Matt J. Rossano, "Artificial Intelligence, Religion, and Community Concern," *Zygon* 36, no. 1 (March 200): 57-75. Laurence Tamatea, "If robots R-US, who am I: Online 'Christian' responses to artificial intelligence," *Culture and Religion* 9, no. 2 (July 2008): 141-160.

[14] Noreen L. Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis: Fortress Press, 2002), 10, ix.

> hold as so important to our human nature that we hope to image them in AI, and how might our choice among these qualities affect any intelligent machine we produce, our relationship with that machine, and our understanding of ourselves?[15]

Herzfeld seeks to develop an understanding of a proper ordering of relationships between humans and the machines we develop and utilize using insights and practices from Christian spirituality. For Herzfeld, spirituality is "the study of the human experience of encounter with an Other or the divine and our lived response to that experience."[16] As artificial intelligences are an "other" to human beings - similar in some ways, yet not the same - this spiritual encounter with that otherness forms the basis of understanding of what we value and seek to image in our creations.

The way that Herzfeld then approaches this set of questions suffers from several problems. First, she explicitly makes her inquiry one regarding "artificial intelligence" and her approaches to understanding the theological and ethical meaning of the human immediately become limited to issues to differing understandings of intelligence, thus ignoring qualities like embodiment and emotion that I believe are crucial to explore (and have been ignored by many AI theorists).[17] Second, while she seeks to address a "human-computer ethic," she shies away from many of the most interesting and challenging ethical questions that arise in the development and utilization of humanoid machines. Third, the way she develops her understanding of the *imago Dei* in the human relies too heavily on the uniqueness of human being's imaging God to take the possibility of robotic

---

[15] Ibid., 5.
[16] Ibid., 8.
[17] Herzfeld actually develops a fairly good critique of the anti-embodiment stance of Kurzweil and Moravec, but this fails to fully translate over into her own understanding of the person.

personhood seriously.[18] Herzfeld does not believe that there can be authentic relationships between humans and machines, that all such relationships are "mere projection and anthropomorphism."[19] She concludes with the idea that, while developing humanoid machines can be a fine thing as long as they are treated non-cavalierly, any humanoid machine and any human relationship with that machine remains subservient to human-human and human-divine relationships, as our "instructions for dealing with the material world are set in the context of, and are subordinate to, instructions regarding proper relationships with other humans."[20] Herzfeld's key ethical question is that of the proper ordering of relationships, so that "given a choice between human relationship or relationship with a computer, we should always choose the human," as though those things were necessarily mutually exclusive and the most pressing ethical question at stake.[21]

Russell Bjork's article, "Artificial Intelligence and the Soul," examines whether "there is an inherent theological conflict between strong artificial intelligence, on the one hand, and biblical teaching regarding the origin of the soul, human worth, and humanity being made in the image of God."[22] As a professor of computer science who also holds a degree in theology, Bjork argues that there is no necessary incompatibility between strong

---

[18] Herzfeld, *In Our Image*, 86-87, 91-93.
[19] Ibid., 91.
[20] Ibid., 92. Interestingly, and perhaps tellingly, Herzfeld as a theologian seeks to develop her understanding by avoiding working in either systematic theology or to "ethical systems and theories" (92), but instead to "spirituality" (8) and the Rule of St. Benedict (92).
[21] Ibid., 93.
[22] Russell C. Bjork, "Artificial Intelligence and the Soul," *Perspectives on Science and Christian Faith* 60, no. 2 (June 2008): 95.

AI and his Christian view of human beings and human relationships with God. Bjork equates the existence of a soul (the "immaterial aspect of humans")[23] with personhood and argues for an emergence-based view of both in an updated view of classical traducianism, concluding that "there would not seem to be - in principle - a *theological* reason why personhood could not emerge in similar fashion from the operation of a sufficiently complex technological artifact."[24] His best argument in favor of the development of strong AI in relationship to human beings comes in his analysis that even if humanoid machines were to be developed in such a way that they included the emergence of a "soul" (and thus were persons, theologically speaking), this would not in any way challenge the nature of humans being made in the image of God. Just as a parent can fully love more than one child, so God can be in a redemptive relationship with more creatures than just human beings.[25] While I would agree with Bjork that a) there is no theological reason to preclude humanoid robots from the concept of personhood and b) that the potential existence of robotic personhood in no way threatens the theological or moral status of human beings as being made in the image of God, Bjork's unfortunate reliance on the equation of personhood with "soul" and lack of theological (much less Christological) basis for understanding personhood leaves many of the problems of some strains of contemporary robotic dualism, which separates the mind / soul from the body and identifies the

[23] Ibid., 97.
[24] Ibid., 98.
[25] Ibid., 100.

"essence" of the person with the former against the latter, firmly in place.[26] There are movements within both robotics and Christian theology to counter this set of assumptions and instead emphasize embodiment and embeddedness in the world as keys to understanding personhood, which will be explored in chapter 4.

Antje Jackelen queries whether, as humans evolve with our technologies, it continues to make sense to describe human beings as "made in the image of God."[27] Her work concentrates on the internalization of technology into human bodies such that "we are becoming "*techno sapiens,*"[28] rather than specifically focusing on the development of humanoid machines, though she utilizes the concept of AI as an analogue, emphasizing the importance of avoiding reductionistic thinking about intelligence and its relationship to human self-concept.[29] Jackelen envisions no theological problem with some forms of cyborgization that are used to heal bodies. She also helpfully acknowledges the importance of global justice issues in the development and utilization of technology.[30] She likewise avoids the problem of simply equating personhood with biological humanity but does not explore it further, theologically or practically.[31] Jackelen's most helpful argument is her critique of thinkers like Kurzweil in arguing against a progressive concept of

[26] This dualism will be analyzed in chapter 2 and its ethical implications explored in chapter 3. In chapters 4 and 5 I will develop and offer an alternative position.
[27] Antje Jackelen, "The Image of God as Techno Sapiens," *Zygon* 37, no. 2 (June 2002): 289.
[28] Ibid., 290.
[29] Ibid., 292.
[30] Ibid., 294. She does not develop these potential issues of justice in any detail here.
[31] Ibid., 296.

evolution.[32] She argues that

> biologically, evolution operates by means of mutation, selection, and adaptation. In terms of biology, there are no values attached to these concepts. Very often, however, and fairly uncritically, we tend to attach values to evolutionary processes. Thus, we talk in terms of adaptation but understand it in terms of improvement. This applies especially to the context of human evolution, in which evolution has been understood as the successive development of consciousness, awareness, language, sense of moral responsibility, understanding of truth, beauty, and goodness, freedom to do evil, and spirituality.[33]

Her addressing of the question of the image of God, however, falls into an unproblematized and technophilic idea of "relational capacity and creativity" along with imperfection in a co-creator-oriented process theological mode of God "luring the world into its eschatological future."[34] Her work here relies strongly on the "created co-creator" model of personhood developed by Philip Hefner, whose work will be explored in depth in chapter 4.

Ian Barbour's work, like Jackelen's, relies on process theology to address the place of artificial intelligence, embodiment, and a unitary (i.e., non-dualistic) view of human persons in a contemporary theological context. His work does not look at the development of humanoid machines beyond using them as an example, but he does consider some models of embodied and sociable robots as his examples. He argues "that it is consistent with neuroscience, computer science, and a theological view of human nature to understand a person as a multi-level psychosomatic unity who is both a biological

[32] This issue will be further discussed in chapters 2&3.
[33] Jackelen, "The Image of God as Techno Sapiens," 296.
[34] Ibid., 299.

organism and a responsible self."[35] His argument that human beings, theologically and scientifically speaking, are best understood as holistic embodied selves and not body-soul dualities that I will utilize in understanding both human and robotic anthropology; he also explores the anthropological role of emotions, sociality, and relation to God.[36] Barbour, like Jackelen, uses an unrevised version of Hefner's created co-creator model of the human person.[37]

Anne Foerst has the interesting distinction of having been the only theologian to hold a position on the faculty of the MIT robotics laboratory. While she no longer works there, she was a theological advisor to their early humanoid robotics projects, Cog and Kismet. In *God in the Machine,* she argues for an understanding of human and robotic persons as embodied and sociable / relational beings and for a future in which there is a healthy community of both human and non-human persons. Ethically, an important part of her argument is that "every empirical criterion that might exclude Kismet from the community of persons will also exclude human beings from it."[38] She argues that exclusivism in anthropology in all of its forms, whether in racism, sexism, or speciesism, damages human beings and our capacity for compassion. Her anthropological focus relies on the importance of embodiment (which I will also use), but also on the existential

---

[35] Ian Barbour, "Neuroscience, Artificial Intelligence, and Human Nature: Theological and Philosophical Reflections," *Zygon* 34, no. 3 (September 1999): 362.
[36] Ibid., 363-364.
[37] Ibid., 364-365.
[38] Anne Foerst. *God in the Machine: What Robots Teach Us About Humanity and God.* (New York: Dutton, 2004), 162.

condition of estrangement and alienation.[39] Her other work on humanoid robotics, "Artificial Intelligence: Walking the Boundary," focused on robotics as an example in current science / religion dialogues.[40] Here she lays out a Tillichian existentialist theology to argue for responsible and engaged dialogue between theologians and scientists regarding the development of artificial intelligence, seeking to avoid problems of both dogmatism and utter relativism, as well as mutual ignorance and apathy, on both sides of the conversation.

The remaining works on theology and robotics address several areas other than anthropology. Robert Geraci traces the apocalyptic tendencies in contemporary robotics and artificial intelligence. He argues that advocates of strong AI take on a form of religious fervor that embodies characteristics of both Jewish and Christian apocalypticism: "alienation within the world, desire for the establishment of a heavenly new world, and the transformation of human beings so that they may live in that world in purified bodies."[41] Michael DeLashmutt likewise examines the functional eschatologies of futurists like Kurzweil and Moravec, particularly their visions for the ends of humanity and their

[39] My own argument regarding the important facets of both human and robotic anthropology will be considered in depth in chapters 4&5, but as I will argue there, the reduction of human essence to just a few aspects, whether freedom, finitude, or alienation, do not do justice to the whole of human life or account for particularity and difference among persons.
[40] Anne Foerst, "Artificial Intelligence: Walking the Boundary," *Zygon* 31, no. 4 (December 1996): 681-693.
[41] Robert M. Geraci, "Apocalyptic AI: Religion and the Promise of Artificial Intelligence," *Journal of the American Academy of Religion* 76, no. 1 (March 2008): 138.

hopes for the future.[42] Matt Rossano argues "that it is precisely the community-building and -strengthening function of religion that will provide its most rational basis for addressing the moral issues associated with the emerging technologies of artificial intelligence, especially those that have a direct impact on brain function."[43] Laurence Tamatea has studied online "Christian" responses to ethical issues and understandings of personhood in emerging AI and robotics literature, stating that the majority of popular responses from openly religious persons have been hostile, envisioning the development of strong AI and humanoid robots as a "Hell" scenario in which humans are "playing God" and the image of God in human beings in threatened.[44]

As I have noted, several of these authors (Herzfeld, Foerst, Barbour, and Jackelen) derive their conceptions of personhood from Hefner's "created co-creators" understanding of theological anthropology, which I believe is insufficient for full consideration of these matters.[45] The "created co-creators" motif takes an abstracted concept of God as Creator, expands such creativity to human beings, and tries to develop a theology of creativity from this, largely in absence of a robust theology of creation's fallenness and redemption in Christ. Without this more robust theological anthropology and centering in the heart of Christian faith, such a motif ends up at best with a notion of human creativity as the

---

[42] Michael W. DeLashmutt, "A Better Life Through Information Technology? The Techno-Theological Eschatology of Posthuman Speculative Science," *Zygon* 41, no. 2 (June 2006): 267-287.
[43] Matt J. Rossano, "Artificial Intelligence, Religion, and Community Concern," *Zygon* 36, no. 1 (March 2001): 58.
[44] Laurence Tamatea, "If robots R-US, who am I: Online 'Christian' responses to artificial intelligence," *Culture and Religion* 9, no. 2 (July 2008): 141-160.
[45] Philip Hefner, *The Human Factor: Evolution, Culture, and Religion*, (Minneapolis: Augsburg Fortress, 1993).

essence and fulfillment of who we are called to be; it tends to baptize any current technological project as "good" that involves human inventiveness, regardless of the cost in other areas and with little regard for the frailties and limitations of human life. As Hefner's work grounds the work of these theologian's reflections on robotics, I will explore his work directly in relating theology to the question of personhood in contemporary robotics. There are some real strengths in his work, which emphasizes the embodiment of human life and our rootedness in our evolutionary and cultural heritage. We are situated in history, and out of that history arises our creativity and relationship to God.

In more conservative forms of theological reflection, technological development is frequently looked upon as a "slippery slope" to a variety of ills, particularly the degradation of human personhood and dignity.[46] This concern about and sometimes rejection of new technologies is frequently coupled with a charge of "playing God," violating inherent human limits, and ignoring the unique sanctity of human life.[47] While not all of the bioconservatives reject all technologies, they are less enthusiastic and far more concerned about the possible abuses and detrimental effects than Hefner and others who follow the created co-creators model.

---

[46] President's Council on Bioethics, *Beyond Therapy: Biotechnology and the Pursuit of Happiness*, (Washington, DC: US Government Printing Office, 2003).
[47] Examples include Bill McKibben, *Enough: Staying Human in an Engineered Age*, (New York: Times Books, 2003), and many of the essays in John F. Kilner, et al., eds., *Genetic Ethics: Do the Ends Justify the Genes?* (Grand Rapids: Eerdmans Publishing, 1997), John F. Kilner et al., eds., *The Reproduction Revolution: A Christian Appraisal of Sexuality, Reproductive Technologies, and the Family*, (Grand Rapids: Eerdmans Publishing), 2000, and C. Ben Mitchell, et al., eds., *Biotechnology and the Human Good*, (Washington, DC: Georgetown University Press, 2007).

One form of this conservative theological anthropology is developed by Brent Waters, who focuses on human finitude and mortality as the framework from which to understand the place of human freedom and the role of technology. Waters takes seriously human embodied life and its limitations, as well as the problems that can arise when we seek to overcome our finitude. Waters rejects dualistic anthropologies in theology as well as technology, focusing on our embodied nature and the relationships to both God and other persons in which we become who and what we are. His theological anthropology will be analyzed, along with Hefner's, in chapter 4. Both the more liberal and more conservative theologies of technology have significant benefits and major flaws, and a new basis of thinking theologically about technology that combines the best insights of both while avoiding their problems must be sought instead. For instance, Hefner emphasizes creativity, relationality, and embodiment without as much consideration of the limitations of human life, while Waters focuses on embodiment and its limits but rejects creativity to adapt the situatedness with which we live. In balancing these perspectives, theology can contribute to a robust understanding of personhood - human or otherwise - that ties into the picture or personhood as embodied, sociable, and situated that arises in some contemporary robotics.

The Theological Idea of the "Person"

A first thing that may be said regarding historical Christian theological thinking about the nature of persons is that it does not, in a substantive sense, begin with human

beings at all. In the Hebrew Bible, the God of Israel is described in a variety of different ways, many of them anthropomorphic. While the Israelite God was worshipped as the Creator of all, beyond the world, and as Spirit rather than flesh, that same God was also "personal," in that God was intimately involved in the lives and history of human beings, first and foremost in the nation of Israel. God heard the cries of the people and delivered them; God spoke through Moses and gave the people the law; God made covenants with the people and keeps them forever; God spoke through the prophets.[48] That is to say, God was not simply a force in the universe, the engine that made all things run. Nor was God simply a local deity, with the same flaws and problems that humans had, only on a greater scale. Rather, it was the claim of the Israelites that the God of the Universe, who created and sustained all things, had personally chosen a people through whom to bless all the nations of the earth. The Jews had been conquered, exiled, returned to the land, and occupied, but they retained the hope that the God who covenanted with them would sustain them and bring about justice on the earth.

Within the milieu of Roman-occupied Palestine, there arose a movement around the person of Jesus, who had preached about the reign of God and had traveled the land feeding and healing the people. He was executed as a traitor, but the movement that had sprung up around him not only remained, but grew. His followers claimed that he was the promised messiah - the anointed one who would bring liberation and redemption to the people of Israel, but that this liberation and redemption existed spiritually, not politically.

[48] Exodus 2:23-24, Exodus 12:31-39, Exodus 20:1-17, Genesis 8-9, Genesis 12-17, Genesis 27-28, Exodus 19-24.

Prayers and praises were offered to God in the name of Jesus, who was understood to have been raised from the dead by God and exalted. He was identified with the Logos of God - the Word, the firstborn of all creation, the one who was one with God.[49]

In the first centuries following the time of Jesus, his followers came to understand him as not only the human messiah who was promised to restore Israel, but as one who was to be worshipped as the exalted one who was united with the God he called Father.[50] The status of Jesus in relation to God was a primary theological problem that divided churches in the third and fourth centuries after Jesus and came to a head as Christianity was legalized within the Roman Empire. There were several problems with leaving the precise nature of Jesus in doubt among the churches: theological, philosophical, and practical. On the one hand, Christians claimed that Jesus was the messiah of the God of Israel who had been called the "Son of God" and had been exalted by God in his resurrection. The God of Israel was One, and the Jewish faith was strictly monotheistic.[51] The Jews had no expectation that the messiah would be anything other than a human being - a great human being, but fully human nonetheless. Jesus lived and died a clearly human life; this life was supernaturally touched by God, giving Jesus the power to heal and cast out demons. Many in the church interpreted his being "Son of God" to be a sonship by adoption - a great human being, even a "divine" being, but not a being who was the same as God. The anti-Nicenes, who held various versions of this position, claimed

---

[49] John 1.
[50] While in my own work and theology, I strongly prefer to utilize gender-neutral terms to refer to God, in this historical section I will use the terms used in this period in order to make the relationships and language as clear as possible.
[51] Deuteronomy 6:4.

that Jesus was "homoiousios" with God the Father - of a similar substance, the firstborn of all creation, and greater than any other human being, but still a creature made by God and elevated to an exalted status.

Christians on the other side of the conflict likewise had to explain the nature of Jesus in relation to the monotheism of Judaism, but they also had to justify both their beliefs and practices surrounding Jesus, particularly the worship of Jesus as the savior. The second commandment of Moses identifies the stakes of the claim: the God of Israel instructs that the people worship no other gods. If Christians were worshipping Jesus, and Jesus was not God, regardless of how exalted by God he may be, they were in clear violation of the commandment. These Christians also claimed Jesus as savior, and only God was in a position to ultimately save the people, to forgive sins, and to conquer death - all of which were being claimed as works of Jesus.[52] Where the Arians claimed that Jesus stood on the "created" side of the Creator / creature distinction, those who became known as orthodox made a unique and controversial claim: that the Jesus they knew and experienced as savior and redeemer was "homoousios" with God the Father - of the same substance. Jesus was not only the Son of God by adoption, but in truth, and was not a created being. Jesus and the Father were One - the same substance - and in claiming this oneness, Christians believed they did not violate the monotheism of the Jews. This position came to be the dominant one within the ecumenical churches, and its opponents were driven out as heretics.

---

[52] Athanasius, "On the Incarnation," *Christology of the Later Fathers*, Edward R. Hardy, ed., (Louisville, KY: Westminster John Knox Press, 1954).

The question remained, however, as to what the exact relationship was between Jesus and the Father. How could they be "One" while Jesus lived a human life, praying to the Father? The Son was claimed by the church to be "begotten, not made," and so was one in substance with the Creator, but did this not make God, in a real way, Two, rather than One? Similar questions were raised with regard to the one known as the Holy Spirit. A resolution was devised by the Cappadocian fathers of the fourth century that was accepted at the Council of Constantinople in 381. Under the Cappadocian resolution, God is One with respect to "ousia" - substance or essence.[53] There is only one God-ness that could or did exist. But, they claimed, God was Three with respect to "hypostasis" - "persons." God was not three beings, so the charge of Tritheism could be avoided; but God was also not simply a single God who appeared in different ways; there really were three persons, three individuated realities, living eternally together as a single God. The persons were distinguishable, but not divisible. In substance, they were identical, but as persons, they were differentiated.

The language of persons is thus utilized derivatively in regard to the lives and beings of humans. Humans are said to be made "in the image of God," and have existences that reflect, in some way, the life and being of God. It is God who is truly, definitively "personal." Any moral or theological sense of the human being as a person must be based on the understanding of the personhood of God. At the Council of Chalcedon in 451, the relationship between the divinity and humanity of Christ in relation to his personhood

[53] Gregory of Nyssa, "An Answer to Ablabius: That We Should Not Think of Saying That There Are Three Gods," *Christology of the Later Fathers*, Edward R. Hardy, ed., (Louisville, KY: Westminster John Knox Press, 1954).

was delineated.[54] Jesus the Christ was claimed as "fully divine and fully human," two natures in a single hypostasis (person). Jesus, who was a single person in two natures, stands as the archetype of the person for human beings, who also exist as distinguishable "persons," although without the characteristic perichoresis of the oneness of God.

I hope to use this theological anthropology to further develop a narrative of the person that is expansive, encompassing a widening circle of grace, yet also inclusive of those who have sometimes been left behind and who are left behind in the technofuturist narrative – the poor and those who do not or cannot embrace the technoscientific prospect. I want to offer a picture of personhood that opens the door for the use of technology within our bodies but subverts and opposes the requirement of "keeping up" by succumbing to market pressures to have technology control every aspect of our lives. This theological picture of personhood is expansive by showing examples of persons beyond the human – our primary example is the Trinitarian divine life itself, understood through the lens of incarnation – though also inclusive of human beings who are left on the margins.

Project Summary

This chapter will provide an overview of current technological trends in the development of humanoid robotics, including the wide variety of applications for which

[54] The Eastern Orthodox, Roman Catholic, and most Protestant churches accept the formulation of the Council of Chalcedon, though a number of other churches, including the contemporary Oriental Orthodox churches, reject the relationship between the natures. All groups except the Assyrian Church of the East, accept the singularity of the hypostasis of Christ.

they are being used (and planned). I will consider these humanoid robots within three broad category areas: 1) human service applications, including healthcare, personal service, and general work replacement robots; 2) military and defense robots, particularly warfighter robots that utilize independent decision-making in both following the laws of armed conflict and particular rules of engagement as well as determining and executing both offensive (shooting) and defensive (running away) maneuvers; and 3) sex and companionship robots that are not used to automate particular work tasks but to substitute for human companionship, including fully humanoid sexbots and robots that are designed to be "friendly" and relational outside of a labor context. This will largely be a technical overview of the breadth of humanoid robot applications currently produced and envisioned. I will also analyze the major funding streams in the development of humanoid robots and the types of applications toward which each stream tends.

Chapter 2 will analyze the works of six prominent authors within contemporary robotic futurism and the particular perspectives on robotic (and human) anthropology within those futurists' writings. This textual analysis of popular scholarly literature is intended to provide a window into the current state of discourse within the field. In exploring these discourses, I hope to gain insight into the implicit and explicit working definitions of personhood within their writings, what they consider the salient characteristics that bring a being either partially or fully into the moral community, what makes someone a "who" rather than a "what." These six authors, Ray Kurzweil, Hans Moravec, Marvin Minsky, J. Storrs Hall, Rodney Brooks, and Cynthia Breazeal, have very different conceptions of personhood and the relative importance of rationality,

intelligence, computation, emotion, embodiment, and relationality within personhood.

Chapter 3 will expand upon the analysis developed in Chapter 2 and will explore some of both the problematic and constructive ethical implications for the development of humanoid robots based on the respective understandings of personhood of the six robotic futurists. I will analyze some of the concrete choices made as a result of these working anthropologies and the directions in which they tend to lead in robotics development in relation to the human moral community, how they both positively and negatively influence the working anthropologies of robotics, and how they can be used constructively in reformulating what a robotic anthropology might look like.

Chapter 4 will explore understandings of personhood in a very different mode, one that begins not with robotics, but with historic Christian theological conceptions of what it means to be a person, human or otherwise. This analysis will begin by considering the theological anthropologies of two contemporary theologians working on emerging technologies - Philip Hefner and Brent Waters. Hefner develops a model of the (human) person as the "created co-creator," which emphasizes human freedom for creative activity in the world, where Waters develops an anthropology focused on the necessity and goodness of human limitations, which I refer to as the "embodied-finite-mortal creature" model. Though they are, in some respects, quite opposite of one another, I argue that they are not entirely incompatible, and that a robust theological anthropology will uphold much of the content of both positions. This chapter will then explore ways in which theological reflection can deepen the understanding of the person developed in contemporary robotics as an emergent mind that is embodied, sociable, and situated.

Chapter 5 will connect the theological analysis of Chapter 4 with robotic futurism in an attempt to begin to develop a constructive theological and ethical narrative of what humanoid robotic personhood could look like in a future that allows for the flourishing of all (human and non-human) persons. Robots will be considered as both potential objects and subjects of moral action and the qualities that contribute to that objectivity and subjectivity. A theological understanding of robotic personhood will finally be explored in the context of particular applications for humanoid robots: in labor replacement and human service, military and defense, and sex and companionship applications.

Applications for Humanoid Robotics

I will begin by looking at a number of the primary areas of current development in humanoid robotics, focusing on three major overlapping themes: labor replacement and human service robots, military and defense robots, and sex and companionship robots. This will not attempt to cover the field of robotics in general but will focus closely on humanoid robots - why they are developed, how they are used, and what constitutes the "humanoid" character of such robots in contrast to traditional robotics. While robots of various kinds have been used in a number of different industries and serve primarily in automation and labor replacement, humanoid robotics represents a new set of applications that cover a broad range of fields.

The first primary area of development consists in applications related to labor replacement and personal services, including healthcare applications. This is a key aspect of humanoid robotics research and development in Japan, as such robots are being

designed to serve an aging society. Robots have traditionally been designed as labor-saving or labor-replacement devices, such as the popular Roomba robotic vacuum, introduced in 2002 by the iRobot Corporation.[55] While the automation of manufacturing is nothing new and stretches back to the invention of the assembly line, robotics development has challenged areas where workers formerly believed their skills, abilities, and decision-making capacities to be safe.[56] Precision robotic surgery is beginning to be used to perform tasks too delicate and detailed for even the most skilled human surgeon; self-driving robots are being shown to be safe over long distances and can be used to replace taxi drivers and over-the-road truckers; and personal care robots can be used to aid in mobility and medical care for senior citizens.[57]

One of the most important areas in which humanoid robotic labor can be used to replace human labor is in dangerous, dirty, or otherwise undesirable labor applications. For instance, robots have been used both to understand the extent of and to clean up the nuclear meltdown at the Fukushima Daiichi plant in Japan.[58] Robots are being developed for use as scouts and rescuers in coal-mining disasters, and they are hoped to be useful in

---

[55] "iRobot Corporation: Our History," http://www.irobot.com/sp.cfm?pageid=203. iRobot was named for the Asimov novel and also builds the PackBot (used primarily in military applications) and the Seaglider (used to clean the Gulf oil spill).
[56] Drew Halley, "A Robot Stole My Job: Automation in the Recession," *Singularity Hub*, December 15, 2010, http://singularityhub.com/2010/12/15/a-robot-stole-my-job-automation-in-the-recession/.
[57] Aaron Saenz, "Robotic Labor Taking Over the World? You Bet - Here Are the Details," *Singularity Hub*, September 12, 2011, http://singularityhub.com/2011/09/12/robotic-labor-taking-over-the-world-you-bet-here-are-the-details/.
[58] Larry Greenemeier, "Robots Arrive at Fukushima Nuclear Site With Unclear Mission," *Scientific American*, March 24, 2011, http://www.scientificamerican.com/article.cfm?id=robots-arrive-fukushima-nuclear.

eventually replacing coal miners in their dangerous jobs.[59] Robots even make the exploration of deep space feasible as they can land on a planet, drive around and record data, and locate traces of water on distant planets.[60] While many of these are fascinating applications in their own right, the ways in which they perform particular human tasks does not say too much about how they can be made fully humanoid.

Some emerging medical applications indicate the directions in which a more fully humanoid human service robot may be developed to do more than simply perform discrete tasks. A "doctor bot" has been envisioned that would fully replace a human physician.[61] The doctor bot would not require the years of training it takes for a human physician to fully develop her skills; the robot's knowledge could be simply downloaded and continually updated with the latest research. But the doctor bot needs more than simple medical knowledge in order to serve as a physician replacement. The doctor bot would also need to be able to talk with patients in an empathetic and understanding way, to perform complex humanoid tasks in medical diagnosis, to make subtle differentiations in symptoms, and to diagnose and treat cases that differed from the "standard" presentation. It is likely that this change will occur over in stages the course of a generation or two, first

---

[59] David Hambling, "Next-Gen Coal Mining Rescue Robot," *Popular Mechanics*, August 23, 2010, http://www.popularmechanics.com/science/energy/coal-oil-gas/next-gen-coal-mining-rescue-robot. Catherine Arnst, "I, Robot: In developing robots to work alongside humans, scientists find even crude facsimiles of human behavior help people accept mechanical colleagues," *Bloomberg Businessweek*, August 20, 2007, http://www.businessweek.com/technology/content/aug2007/tc20070818_479361.htm.
[60] NASA Jet Propulsion Laboratory, http://www-robotics.jpl.nasa.gov/applications/index.cfm.
[61] Alison Diana, "12 Advances in Medical Robotics," *Information Week Healthcare*, January 29, 2011, http://www.informationweek.com/news/galleries/healthcare/patient/229100383.

under present conditions in which vast medical databases are available to physicians in the course of their everyday clinical practice, secondly to robots who assist physicians in their diagnosis and treatment by interacting directly with the patients, then to the deprofessionalization of medical practitioners who assist the robots, and finally to the full replacement of human physicians with robots. The replacement of other human services, such as nursing, are at a more advanced stage of development now, as robots are designed who can assess the needs of patients, lift them, and perform simple daily tasks.

A second critical area of humanoid robotics development in the United States and elsewhere is that of military and defense robots. These robots take a number of different forms and serve in both offensive and defensive capacities. Peter W. Singer has analyzed the automation of the United States military in *Wired for War*, currently the most comprehensive treatment of the subject of emerging military technologies available from non-classified sources.[62] One of the key goals of current developments in military technology is to remove humans (and thus, human casualties) from the battlefield.[63] Singer's analysis begins with looking at the ways in which many military functions have been made more remote, as when soldiers stationed in a bunker in Nevada fight remotely on battlefields in Iraq and Afghanistan using interfaces that are designed to resemble video games.[64] Singer also highlights current developments in drone warfare and the human-less

[62] P. W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, (New York: Penguin, 2009), 14.
[63] Ibid., 34.
[64] Ibid., 32-35, 367.

surveillance and fighting machines that are being used across the world.[65] Next he considers the animalistic robots that the US military is using for applications from carrying machinery to defusing landmines to serving alongside human warfighters.[66] Finally he looks into the future of warfare and the development of fully humanoid, autonomous warfighter robots that can be used to replace soldiers in ground combat and can make decisions on their own, for instance, when and whether to shoot.[67]

Georgia Tech roboticist Ronald Arkin has begun to explore ways in which humanoid warfighter robots could be programmed to be more ethical on the 21st century battlefield than human soldiers.[68] On the one hand, humanoid robot warfighters require extensive programming in order to follow the laws of war and rules of engagement, and there are issues with the removal of human choice from offensive strategic and tactical decisions in war. On the other hand, it is very easy for robot warfighters to avoid many of the classic revenge- and power-oriented problems associated with humans in war, such as rape, pillage, and the intentional targeting of civilians.[69] Robot fighters do not tire (except when they need recharging), do not require food and water, never experience PTSD, and never sleep. He proposes a set of solutions involving an iterative process of calculation / rules / action / reflection in order to develop ethical decision-making in autonomous

[65] Ibid., 36.
[66] Ibid., 20, 23.
[67] Ibid., 126.
[68] Ronald Arkin, *Governing Lethal Behavior in Autonomous Robots*, (Boca Raton: CRC Press, 2009).
[69] Ibid., 29-36.

military robots.[70]

A third key avenue for humanoid robotics development encompasses a broad range of sex and companionship robots, from proto-humanoid robotic pets to sexbots of various capabilities. In *Love and Sex with Robots*, David Levy describes the disparate variety of ways in which humanoid robots are being developed and utilized for applications to fulfill the love, sex, and companionship needs of human beings.[71] The companionship applications of humanoid robots are as varied as the human beings for whom they are developed. He looks at the relationship between biological pets and the emotional connections that people develop with electronic devices, indicating that, at some level, they may become comparable both in terms of the relationships and projections of emotion that people place on both pets and robots.[72] The new field of robotherapy joins medical, human service, and companion robots as it "concentrates on the task of employing interactive robots as therapeutic companions for people who have psychological problems or are handicapped physically, emotionally, or cognitively."[73] As several recent documentary and fictional films have shown that some humans can develop projectional emotional relationships with human-looking (but inanimate) dolls, humanoid robot companions serve as the next realistic phase of those types of connections.[74] Humanoid robots can produce greater levels of interactivity with their owners, including

---

[70] Ibid., 108-113.
[71] David Levy, *Love and Sex with Robots: The Evolution of Human-Robot Relationships*, (New York: Harper Perennial, 2008).
[72] Ibid., 14, 64-104.
[73] Ibid., 9-10.
[74] These films include Craig Gillespie's fictional *Lars and the Real Girl* (2007) and documentaries like Allison de Fren's *The Mechanical Bride* (2010).

sexual activity. Sexual technology (e.g., vibrators and dildos) and paid sexual partnership are already common features of human sexuality, and robotic sexuality combines them in ever-more-lifelike forms. Unlike prostitution, relationships with robotic sex partners are presumably legal and can provide sexual options for fetishes that humans may not otherwise be able to fulfill. Robotic sex partners are, by design, always willing partners, and may be available without the "messy" realities and needs of human partners. While this raises a variety of interesting ethical questions, the broad availability and acceptance of sex with robots will likely soon be a regular feature of human sexual variety.[75]

The humanoid character of these robots can be differentiated into two broad, if overlapping, types, which I would designate as sexbots and companionbots. Both types can be used for both sexual and companionship purposes, but the ways in which they interact with humans would be very different. On the one hand, those robots designated as "sexbots" would exist solely to serve the needs of their human "owners" and would be programmed to be slavish in their responses. They would not elicit empathy from the humans who interact with them and would be designed strictly to do as they were told, requiring nothing in return. Sexbots, like current sex dolls, offer a form of intimacy and companionship without responsibility toward the robot. They would provide an escape from the ethical and relational responsibilities of real human companionship. Companionbots, on the other hand, would have "needs" of their own and would be

[75] Many of these ethical questions surround problematic, dangerous, and otherwise unacceptable sexual practices. For instance, is it better for pedophiles to be able to act out their desires on childlike sex robots or not? Does it increase or decrease the likelihood of their acting on actual children? What about rape fetishists? Sadists?

designed to elicit empathy from the humans who interact with them. Instead of providing an escape from human companionship, they would provide a form of real companionship and relationship. They would exist in a more reciprocal relationship with their human companions, whether real or perceived, but in either case, the intent behind their design would be different. For instance, the sexbots currently being designed are intended to be slavish, to not have interests, needs, or wills of their own, and to provide a sexual outlet for those who have difficulty with or choose not to enter into sexual relationships with human beings.[76] They are specifically designed to be different from "real women" in key ways, and thus offer a less-than-human "relationship." But they could instead be designed to have needs, interests, and choices of their own, to require care from the humans with whom they interact, and to more fully mimic real human-human relationships.

Policy and Funding of Humanoid Robotics Development

Humanoid robotics are being developed around the world by a wide range of parties, and this section will focus on the primary funding sources and policies behind that development.

In the United States, there is substantial federal funding available for robotics development, primarily through the National Science Foundation's (NSF) new (2011) National Robotics Initiative (NRI), a collaborative effort between three agencies: the National Institutes of Health (NIH), the Department of Agriculture (DoA), and the

[76] *Sex Robot*, Discovery Fit and Health, 2011.

National Aeronautics and Space Administration (NASA).[77] The funding for NASA also includes grants provided for research by the Department of Defense (DoD). The NRI focuses on funding primary research in robotics, "fundamental research and education by academia and industry built on open platforms, enabling demonstration systems and transfer to commercial exploitation."[78] It receives $40-50 million per year in NSF grants[79] and another $40 million through the DoD's Defense University Research Instrumentation Program (DURIP), which is specifically designated for the development of robotics for military applications.[80] These are recent additions to public funding at the federal level, which add to a variety of autonomous systems being developed through the Defense Advanced Research Projects Agency (DARPA), with an annual unclassified budget of $3.2 billion,[81] the National Institute of Standards and Technology (NIST),[82] and the Department of Homeland Security (DHS).[83]

Globally, public funding for humanoid robotics research is rapidly growing. According to a report commissioned by the US Robotics corporation:

---

[77] National Robotics Initiative, http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503641&org=CISE.
[78] National Robotics Initiative, http://www.nsf.gov/pubs/2011/nsf11553/nsf11553.htm.
[79] National Robotics Initiative, http://www.nsf.gov/pubs/2011/nsf11553/nsf11553.htm.
[80] Tom Kalil and Chuck Thorpe, White House Office of Science and Technology Policy, "Supporting the President's National Robotics Initiative," August 3, 2011, http://www.whitehouse.gov/blog/2011/08/03/supporting-president-s-national-robotics-initiative. "Program Announcement: Department of Defense (DoD) Fiscal Year 2012 Defense University Research Instrumentation Program (DURIP)," http://www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2011/BAA-AFOSR-DURIP-2011.ashx.
[81] US DARPA, http://www.darpa.mil/NewsEvents/Budget.aspx.
[82] US National Institute of Standards and Technology, http://www.nist.gov/el/isd/robotics.cfm.
[83] US National Institute of Standards and Technology, http://www.nist.gov/el/isd/ks/response_robot_test_methods.cfm.

> The European Commission recently launched a program through which 600 mill Euros are invested in robotics and cognitive systems with a view to strengthen the industry, particularly in manufacturing and services. Korea has launched a comparable program as part of their 21st century frontier initiative, committing to invest $1B in robotics technology over a period of 10 years. Similar, but smaller programs are also in place in Australia, Singapore, and China. In the United States, funding has been committed for unmanned systems within the defense industry, but very few programs have been established in the commercial, healthcare, and industrial sectors.[84]

While the US investment in non-military technologies has begun to grow in 2011, the vast majority of US public investment in robotics remains driven by military concerns, while countries like South Korea and Japan have invested primarily in human service applications.

Private investment in robotics varies globally as well, with many of the powerhouse companies in the United States and Japan. Corporations based in the United States, like iRobot and Foster-Miller, have emphasized military and home-cleaning applications.[85] In Japan, by contrast, automotive manufacturers Toyota and Honda have led the way in developing robots that can walk, talk, and otherwise interact very much like humans. Honda's ASIMO (an acronym for Advanced Step in Innovative Mobility, but also a play on science fiction author Isaac Asimov's name) was introduced in 2000 and was one of the first widely accessible and affordable bipedal and programmable robots.[86] Its engineering and infrastructure have undergone several revisions during the past decade and it is now able to effectively navigate complex terrain, respond to simple

[84] US Robotics, "A Roadmap for US Robotics - From Internet to Robotics," May 21, 2009, http://www.us-robotics.us/reports/CCC%20Report.pdf.
[85] Singer, *Wired for War*, 21-32.
[86] Honda Corporation, "History of ASIMO," http://asimo.honda.com/asimo-history/.

voice commands, and conduct an orchestra.[87] Toyota recently debuted its new healthcare robots that can assist patients with balancing and walking, and can lift a person who has difficulty with mobility for movement between locations.[88]

The differences between research areas being funded by both the public and corporations in different countries is stark in what they value and where they envision the future of technological development in robotics. The different aspects of human life that are incorporated into these robots and how these applications are addressed will be a central focus in chapter 3.

[87] Ibid.
[88] Evan Ackerman, "Toyota's Healthcare Robots Are Ready to Help You With Absolutely Everything," *IEEE Spectrum*, November 4, 2011. Toyota Motor Corporation Press Release, "TMC Shows New Nursing and Healthcare Robots in Tokyo," November 1, 2011, http://pressroom.toyota.com/releases/tmc+shows+new+nursing+healthcare+robots+tokyo.htm.

Chapter 2: Robotic Futurism and the Idea of the "Person"

September 2025: I am awakened this morning promptly at 6:30 by the HAnA 2.5 (Humanoid Analogue Assistant) robot I call Colleen. Colleen informs me of the morning's important new items as I get out of bed, and she recommends an egg white and vegetable omelet for a healthy breakfast. I smell the food that she is cooking as I get dressed, and as soon as I finish breakfast, it's time to head off to work. I put the finishing touches on my morning lecture as Colleen drives to the campus and parks the car. She reminds me that I have a meeting with a student after class at 9:30 and offers to get me a cup of tea. As she walks down the stairs to the lounge, I wonder how I ever got along without her. She walks on two legs and takes care of most of my mundane administrative and households tasks. Her face reminds me that she's a robot - big eyes and wide, smiling lips assure me that she's friendly and trustworthy. I interact with her almost like she's another human being, except that I don't have to feel badly about asking her to do things for me; that's why she was built and why I bought her. She's not exactly a fascinating conversationalist - she is full of information but has few insights on how I should think about the article I'm working on - but that's true of most people, I suppose. She's better than I am at tennis and doesn't need to do yoga to maintain her balance and flexibility. While I'm at work today, I will have her head home and clean up the kitchen and dining room before our friends come over for dinner tonight. They have the latest HAnA, which has better fine motor skills and a superior natural language processor; there's never a lag time in verbal responses like there can be with the 2.5. I find myself being polite to Colleen, always saying "please" and "thank you" just like I would to a human being, but I catch myself

sometimes. She's just a robot, after all.

This chapter will analyze major trends in contemporary robotic futurism and the underlying anthropologies and ethics envisioned by robotic futurists. I will explore what people in the field of robotics understand to make a machine like a person and why. Is Colleen a person? What would make her so? Is it her intelligence, her body, her relationships, her learning? The theorists whose work will be analyzed are technical experts who have been developing robotic and related technologies for decades. They have designed and built their own robots, trained current developers of robotics, written popular and scholarly literature on the development of robotics, and serve in key consulting capacities for major funders of robotics in the US and elsewhere. They are all highly influential thought leaders and are the major voices shaping popular and technical discourse on the future of robotics, so that even when they speak outside their areas of expertise (e.g., in philosophy), they must be taken seriously and responded to thoughtfully because of their outsized influence. The early generations of robotic futurists have traditionally been older, scientifically well-educated, Western white men with a bent toward libertarian-capitalist politics. Several have claimed a neo-Cartesian, materialist-dualistic anthropology and an understanding of evolution as "progress" from relative simplicity toward complexity. They have each been visionary technical thinkers, but there are numerous social, political, economic, and ethical reasons to question their judgment as they proclaim an inevitable technological future. On the other hand, there is reason for hopefulness, as the current generation of researchers begins to move beyond this early

vision into a more complex and embodied understanding of both robots and humans.

A variety of major theorists' contributions to the field will be considered. While all of them are still writing, they can be roughly divided into three chronological groups, which I will designate respectively as first-, second-, and third-wave.[89] These designations correspond approximately to the time of their formal training in robotics. The first-wave group began their training in the 1960s and 1970s, the second-wave in the later 1970s and 1980s, and the third-wave in the 1990s to the present. These "generational" designations also correspond to key modes of thought and understandings of robotic and human personhood and development that influence their work. Kurzweil, Moravec, and Minsky represent the first-wave and tend to focus highly on computational development in their humanoid robotics. The first-wave thinkers tend to focus on hopefulness for the future, even as robots take over most human functions. In the second wave, Brooks' and Hall's work, and to an extent, Minsky's,[90] focus on embodiment, emotional development, and interaction with human beings beyond mere computational capacity; they are both more ambivalent about the impact on the future of human life, indicating that it is likely not the unalloyed blessing that Kurzweil and Moravec envision, but also not the dystopic hell of American science fiction. The third-wave roboticsts are represented in the work of

[89] These designations of the respective "waves" are my own, and I have not seen robotic futurists so designated elsewhere, though other writers have made distinctions between groups and their respective patterns of thought. These have sometimes been associated with various university programs: The MIT school, the Carnegie-Mellon school, etc., though I find that generational comparisons track more closely to the distinctions I am choosing to highlight here.

[90] Minsky is generationally almost proto-first-wave, but some of his recent work on emotion fits more with the second-wave.

Cynthia Breazeal, whose work focuses even more closely on the embodied, social lives of robots and humans together. In her work, sheer increases in processing power are not evidence of greater humanity in robots; rather, the more that robots interact with people in ways that seem "alive" - smiling, care, loneliness, talking - the more they are considered to be like people.

In analyzing the work of these robotic futurists, I will examine their respective visions of human and robotic persons and their visions of the future of both humans and machines. I will develop an analysis of their thought and work through the three "waves." Some of the ethical implications of their anthropologies and visions of the future will be explored in chapter 3.

Ray Kurzweil

Ray Kurzweil is an inventor and futurist whose early work involved developing computer-generated music and text-to-speech programs for the blind. He studied at MIT in the late 1960s under the tutelage of Marvin Minsky, another influential roboticist whose work will be discussed later. His work in technological futurism has been widely popularized, including several books, educational institutions and political advocacy, and two documentary films. His written work follows two related tracks: one on the future of robotics, and one on radical life extension. His work on robotics began in 1990 with the publication of *The Age of Intelligent Machines*,[91] in which Kurzweil forecasted future developments in artificial intelligence. This was followed up in 1999 with *The Age of*

[91] Ray Kurzweil, *The Age of Intelligent Machines*, (Cambridge, MA: MIT Press, 1990).

Spiritual Machines,[92] in which he developed his distinctive idea of the "Law of Accelerating Returns," (LoAR) which argues for an ever-increasing rate of technological change leading to an inevitable post-biological future. His most recent book on the future of robotics is *The Singularity is Near: When Humans Transcend Biology*, published in 2005, in which he marries the idea of the accelerating evolution of technology to the concept of the Singularity, a term coined by Vernor Vinge in 1993 that refers to the point at which the rate of technological (and thus cultural) change is so rapid that the human mind cannot keep up and beyond which the future world cannot be imagined.[93] His work on radical life extension has incorporated nutrition and exercise information, recommendations for a huge number of vitamin and mineral supplements, and hope for radical developments in anti-aging technologies, including nanoscale manipulation of cells. These ideas have appeared in three books, *The 10% Solution for a Healthy Life*, *Fantastic Voyage: Live Long Enough to Live Forever*, and *Transcend: Nine Steps to Living Well Forever*.[94] His stated hope is to live long enough to be able to take full advantage of the benefits he sees coming from advanced AI and life extension technologies, specifically to witness the Singularity and to avoid death until his mind can be uploaded into a more

---

[92] Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, (New York: Viking, 1999).
[93] Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology*, (New York: Viking, 2005). Vernor Vinge, "The Coming Technological Singularity: How to Survive in the Post-Human Era," VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, March 30-31, 1993.
[94] Raymond Kurzweil, *The 10% Solution for a Healthy Life: How to Reduce Fat in Your Diet and Eliminate Virtually All Risk of Heart Disease and Cancer*, (New York: Three Rivers Press, 1994). Ray Kurzweil and Terry Grossman, *Fantastic Voyage: Live Long Enough to Live Forever*, (New York: Plume, 2005). Ray Kurzweil and Terry Grossman, *Transcend: Nine Steps to Living Well Forever*, (New York: Rodale, 2010).

permanent substrate.[95]

Kurzweil's ideas have found a foothold in the high-tech community, and he has started the Singularity University, an educational nonprofit which includes partners like Google, Nokia, and Cicso, as well as university sponsors,[96] and the Singularity Institute for Artificial Intelligence, which serves as a key think tank and policy organization in favor of advancing toward a Singularity. As Kurzweil explains, "The key idea underlying the impending Singularity is that the pace of change of our human-created technology is accelerating and its powers are expanding at an exponential pace."[97]

The anthropology underlying Kurzweil's work is a neo-Cartesian materialist patternism, in which a person is not to be identified primarily with his or her body, nor with a soul, but as "a pattern of matter and energy that persists over time."[98] A patternist, in his view, is "someone who views patterns of information as the fundamental reality."[99] In this interpretation, reality is not a combination of matter and energy; instead, the fundamental makeup of the universe is the information contained in various substrates, including human persons. He believes that "it is not demeaning to regard a person as a profound pattern (a form of knowledge), which is lost when he or she dies. That, at least, is the case today, since we do not yet have the means to access and back up this

---

[95] *Transcendent Man: The Life and Ideas of Ray Kurzweil*, directed by Barry Ptolemy, William Morris Endeavor, 2009.
[96] http://singularityu.org/about/partners/.
[97] Kurzweil, *The Singularity is Near*, 7-8.
[98] Ibid., 383.
[99] Ibid., 5.

knowledge."[100] The substrate itself is unimportant, and in the case of human beings, in desperate need of replacement, as the mortal body allows the pattern to be lost forever when the person dies. Kurzweil says of himself, "I am principally a pattern that persists in time. I am an evolving pattern, and I can influence the evolution of my pattern. Knowledge is a pattern, as distinguished from mere information, and losing knowledge is a profound loss. Thus, losing a person is the ultimate loss."[101] It is his great hope that advanced AI capabilities will be combined with advanced in neuroscience to make it possible for a person to upload the contents of his or her brain into a computer in order to create a form of immortality, because "it is the persistence and power of patterns that support life and intelligence. The pattern is far more important than the material stuff that constitutes it."[102] His Cartesianism is evident in the form and function of his skepticism about the material world: "I don't know for sure that anything exists other than my own thoughts."[103]

For Kurzweil, humans are patterns of information, just as the world is a very complex pattern of information, and this is reflected in his understanding of the purpose of human life, and the purposiveness of the universe itself. "In my view, the purpose of life – and of our lives – is to create and appreciate ever-greater knowledge, to move toward greater 'order'…the purpose of the universe reflects the same purpose as our lives: to

[100] Ibid., 372.
[101] Ibid., 385-6.
[102] Ibid., 388.
[103] Ibid., 390.

move toward greater intelligence and knowledge."[104] Intelligence is both the purpose and goal of the existence of the universe, and both Kurzweil's understanding of persons and his understanding of evolution and evolutionary teleology reflect this belief.

The founding principle upon which Singularitarians base their movement is what Kurzweil calls "the Law of Accelerating Returns." This "Law" is based upon Moore's Law, in which the number of transistors that can be placed on an integrated circuit doubles roughly every two years while maintaining a similar cost per circuit – an exponential increase over time. While Moore's Law has largely held to be true over the course of the past 45 years, Kurzweil's formulation takes this technological trend in processing power and expands its scope to cover technological evolution in general throughout the course of history. The basic form of the LoAR claims that there is an "inherent acceleration in the rate of evolution, with technological evolution as a continuation of biological evolution."[105] This "Law" is then the cause of the coming Singularity, as "the ongoing acceleration of technology is the implication and inevitable result of what I call the law of accelerating returns, which describes the acceleration of the pace of and the exponential growth of the products of an evolutionary process… The Singularity is the inexorable result of the law of accelerating returns."[106]

Kurzweil believes that evolution has several features, each of which plays out in technological evolution: evolution is exponentially progressive and teleological, it is oriented toward knowledge, complexity, and order, and it is inevitable along a particular

[104] Ibid., 372.
[105] Ibid., 7.
[106] Ibid., 35-36.

trajectory, specifically an intellocentric and technocentric trajectory. This applies whether one is discussing biological evolution, cultural evolution, or any other kind of iterative process of change.

First, evolution is exponentially progressive and teleological; this is the basis for the "Law of Accelerating Returns." Kurzweil's argument runs thusly:

> - An evolutionary process is not a closed system; therefore, evolution draws upon the chaos in the larger system in which it takes place for its options for diversity; and
> - Evolution builds upon its own increasing order.
> Therefore:
> - In a evolutionary process, order increases exponentially.
> Therefore:
> - Time exponentially speeds up.
> Therefore:
> - The returns (that is, the valuable products of the process) accelerate.[107]

Although "evolution" here is a fully natural process, it has a specific trajectory along which it inherently runs. The "returns" are exponential, starting off slowly and building until they seem almost infinite. "The first idea is that human progress is exponential (that is, it expands by repeatedly multiplying by a constant) rather than linear (that is, expanding by repeatedly adding a constant). The second is that exponential growth is seductive, starting out slowly and virtually unnoticeably, but beyond the knee of the curve it turns explosive and profoundly transformative."[108] According to Kurzweil's reading of evolution, "exponential growth is a feature of any evolutionary process, of which technology is a primary example."[109] While it is often the case that newly evolved life forms do experience exponential growth in a suitable habitat, there are

[107] Kurzweil, *The Age of Spiritual Machines*, 32-33.
[108] Ibid., 10.
[109] Kurzweil, *The Singularity is Near*, 12.

natural limitations of resources that prevent exponential growth from continuing forever, whether in the form of food shortage, predators, or other limits. Likewise, the exponential growth that does occur is not a value-laden process leading to "returns" of intelligence or progress.

For Kurzweil, evolution is not simply a process by which things change and new life forms are adapted to particular habitats. Evolution is oriented toward developing order and information out of chaos. Though this runs counter to understandings of evolution typically held by evolutionary biologists, Kurzweil defends it adamantly as the way of the universe, claiming that "evolution is a process of creating patterns of increasing order."[110] The course of the evolution of atoms, molecules, cells, and earthly life leads toward the development and utilization of information, which (for Kurzweil) is the basic form of "stuff" in the universe. The universe is made of information, and so everything drives toward the increase of information and order. "Evolution increases order, which may or may not increase complexity (but usually does). A primary reason that evolution – of life-forms or of technology – speeds up is that it builds on its own increasing order, with ever more sophisticated means of recording and manipulating information."[111] Kurzweil claims that limitations that are found in the natural world that prevent a particular species from continuing exponential growth without end are circumvented in this theory of evolution when it comes to information and technology. "The two resources it [evolution] needs – the growing order of the evolving technology itself and the

[110] Ibid., 14.
[111] Ibid., 40.

chaos from which an evolutionary process draws its options for further diversity – are unbounded."[112]

Evolution's trajectory is not only toward the development of order, it is directed toward intelligence and technology, with humanity standing at the pivotal stage of progress to birth the Singularity and the full consciousness of the universe. He claims that "the introduction of technology on Earth is not merely the private affair of one of the Earth's innumerable species. It is a pivotal event in the history of the planet. Evolution's grandest creation – human intelligence – is providing the means for the next stage of evolution, which is technology."[113] Kurzweil's reading of evolution makes himself and his technological creations the pinnacles of history, claiming that "evolution has been seen as a billion-year drama that led inexorably to its grandest creation: human intelligence."[114] This evolutionary trajectory is as inevitable as it is directed toward intelligence. There is no way to stop it, change it, or consider it in another way. "Once life takes hold on a planet, we can consider the emergence of technology as inevitable."[115] Kurzweil claims that this process of increasing intelligence is "inherent," "inevitable," and can only be understood, not changed. "The Singularity denotes an event that will take place in the material world, the *inevitable* next step in the evolutionary process that started with biological evolution and has extended through human-directed technological evolution."[116]

In Kurzweil's perspective, the Singularity thus represents the inevitable fate of

---

[112] Kurzweil, *The Age of Spiritual Machines*, 35.
[113] Ibid., 35.
[114] Ibid., 5.
[115] Ibid., 17.
[116] Kurzweil, *The Singularity is Near*, 387, emphasis added.

human beings and the universe, in which the pace of technological change "will still be finite but so extreme that the changes they bring will appear to rupture the fabric of human history…The Singularity will represent the culmination of the merger of our biological thinking and existence with our technology."[117] This will be disruptive, to be sure, but it will not be a bad thing for human beings: "The Singularity will allow us to transcend these limitations of our biological bodies and brains. We will gain power over our fates. Our mortality will be in our own hands. We will be able to live as long as we want…We will fully understand human thinking and will vastly extend and expand its reach."[118] This utopian view of the future gives humans unprecedented power. He claims that greater intelligence also provides greater goodness, as "our technology will match and then vastly exceed the refinement and suppleness of what we regard as the best of human traits."[119]

Not only will we be better as people in this newly intelligent universe, Kurzweil claims, but the political and social problems that have always plagued humanity will be overcome, as they are fundamentally problems of ignorance, rather than a lack of political will to fix them. "Emerging technologies will provide the means of providing and storing clean and renewable energy, removing toxins and pathogens from our bodies and the environment, and providing the knowledge and wealth to overcome hunger and poverty."[120] Power disparities and economic inequality do not need to be addressed in the present because technology will solve the problems of inequality by making everything

[117] Ibid., 9.
[118] Ibid., 9.
[119] Ibid., 9.
[120] Ibid., 371-2.

cheap and everyone wealthy: "These technologies will create extraordinary wealth, thereby overcoming poverty and enabling us to provide for all of our material needs by transforming inexpensive raw materials and information into any type of product."[121] Addressing the

> possibility that through these technologies the rich may gain certain advantages and opportunities to which the rest of humankind does not have access. Such inequality, of course, would be nothing new, but with regard to this issue the law of accelerating returns has an important and beneficial impact. Because of the ongoing exponential growth of price-performance, all of these technologies quickly become so inexpensive as to become almost free.[122]

Kurzweil's vision of future persons - in which he expects to be included - depicts lives of pleasure, intelligence, and unbounded experiences that transcend the limits of mortality and human embodiment. The body is here only to be discarded bit by bit, replaced by more durable parts.

> We've eliminated the heart, lungs, red and white blood cells, platelets, pancreas, thyroid and all the hormone-producing organs, kidneys, bladder, liver, lower esophagus, stomach, small intestines, large intestines, and bowel…We will not notice the absence of many of our organs, such as the liver and pancreas, since we do not directly experience their operation. But the skin, which includes our primary and secondary sex organs, may prove to be an organ we will actually want to keep, or we may at least want to maintain its vital functions of communication and pleasure.[123]

He is more than happy to get rid of anything that he does not consider to be intrinsic to his understanding of himself – all the messy parts of biological life as we know it. But he does make it clear what he values. By noting that the skin system includes the primary

[121] Ibid., 396-7.
[122] Ibid., 469.
[123] Ibid., 307.

and secondary sex organs, he makes it clear that he is speaking to and about men. Men carry their sex organs on the exterior of their bodies, while women's primary sex organs cannot be understood as part of the skin in any reasonable sense. He rejects all of the parts of his body that he sees as secondary, not directly connected to his experiences of the world. But at the same time he wants to remain active and virile, and sexuality is one of the few bodily pursuits he finds valuable.

Kurzweil's world is one in which humanity has been transcended at the same time that it becomes immortal. The universe wakes up to its own intelligence and an ever-increasing order is pursued naturally and without interruption. Resources will be abundant, everyone will be wealthy, and the problems that have plagued humanity throughout its existence will be left to the past. Despite the development of most advanced AIs by the military, war will not be a problem in the future, and the power of the market will lead to freedom, not constraint.[124] His vision is closely aligned with that of his peer, Hans Moravec, to whose work I will now turn.

Hans Moravec

Hans Moravec is the founder of Carnegie Mellon University's influential robotics program. He came of age in the robotics work of the 1970s at Stanford, where he did his graduate study and began to not only develop functional mobile robots but to dream of

[124] Ibid., 333.

their future as the next phase of intelligent life in the universe.[125] His first major book on the future of robotics was *Mind Children*, published in 1988, which proclaimed robots as the future of humanity's evolution and intelligence in the universe. At a time before most Americans had computers in their own homes, Moravec predicted the development of a "postbiological" intelligence, "a world in which the human race has been swept away by the tide of cultural change, usurped by its own artificial progeny."[126] This work was advanced further a decade later in *Robot: Mere Machine to Transcendent Mind*, a 1999 work that claimed that robots would achieve human level intelligence by 2040 and will have far surpassed humanity within 10 years after that.[127] Since 2003 he has worked on developing mobile utility robots at the Seegrid Corporation in addition to his teaching duties.[128]

In his earlier work, Moravec envisions the future of humanity as not belonging to humanity (or our biological descendants) at all, but to robots as the "mind children" of human beings. Moravec envisions that, "unleashed from the plodding pace of biological evolution, the children of our minds will be free to grow to confront the universe and fundamental challenges in the larger universe. We humans will benefit for a time from their labors, but sooner or later, like natural children, they will seek their fortunes while we,

---

[125] Hans Moravec, *Robot: Mere Machine to Transcendent Mind*, (New York: Oxford, 1999), vii.
[126] Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, (Cambridge, MA: Harvard, 1988), 1.
[127] Moravec, *Robot*.
[128] Hans Moravec, Official CV, http://www.frc.ri.cmu.edu/~hpm/hpm.cv.html.

their aged parents, silently fade away."[129] A few of us might remain as appreciated, though unintelligent, ancestors, or we may (if we are fortunate) be allowed to live on in perpetuity by uploading the functional contents of our brains to computers to be merged with the greater robotic intelligence.

His vision parallels (but does not directly utilize) Kurzweil's understanding of technology as an inevitable and ever-accelerating process. Moravec does not see the coming robotic revolution as one of a Singularity, a point in history past which human beings will not be able to see what lies beyond, for we will likely not be present at all to keep up with the pace of change. Rather, his dominant metaphor is one of ancestry and descent, with occasional metaphors of market competition included as well. His understanding is that natural selection will soon select against us, and we will be surpassed by our own devices. "Such machines could carry on our cultural evolution, including their own construction and increasingly rapid self-improvement, without us, and without the genes that built us. When that happens, our DNA will find itself out of a job, having lost the evolutionary race to a new kind of competition."[130]

While Moravec's vision is one that many would deem apocalyptic,[131] he finds it to be a hopeful one. The end of humanity is one in which parents happily see their children's accomplishments surpass their own. But he also holds out some hope for a different kind of immortality through these machines. He claims that "it is not necessary to adopt a

---

[129] Moravec, *Mind Children*, 1.
[130] Ibid., 2.
[131] Robert Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*, (New York: Oxford University Press, 2010).

mystical or religious stance to accept the possibility" of immortality.[132] Rather, in order to achieve existence without end, "a computation in progress - what we can reasonably call a computer's thought process - can be halted midstep and transferred, as program and data read out of a machine's memory, into a physically different computer, there to resume as though nothing has happened…a human mind might be freed from its brain in some analogous (if more technically challenging) way."[133] We might indeed live on by the transferring of our neural processes to a more flexible and durable substrate, in Moravec's view, thereby gaining plasticity, the ability to direct our own mental growth and development endlessly and without the constraints of bodies that pass away. This idea of "transmigration" would likely involve an advanced form of brain mapping that striped away progressive layers of brain matter and reencoded them onto a new computer substrate until the physical brain was gone. The body would die and the mind would wake up "to a shiny new body of the style, color, and material of your choice. Your metamorphosis is complete."[134]

Moravec, like Kurzweil, rejects what he calls the "body-identity position" in favor of a "pattern-identity" understanding of what makes someone the person that she is.[135] The body-identity view is one in which the person's identity as the individual that she is is intimately connected to the particular physical body in which her life takes place. If one retains the same mental structure (assuming such a thing was possible), but takes

[132] Moravec, *Mind Children*, 4.
[133] Ibid., 4.
[134] Ibid., 110.
[135] Ibid., 116.

on a different body, one is, in a significant way, a different person. The new person may have some of the same thoughts and memories as the old person, but there is a tangible shift in identity that comes with the transfer of media. In contrast, the pattern-identity view, which both Moravec and Kurzweil support, "defines the essence of the person…as the pattern and process going on in my head and body, not the machinery supporting that process. If the process is preserved, I am preserved. The rest is mere jelly."[136] He defends this position based on the fact that the particular atoms and cells that comprise the body change over the course of one's life, so one's identity cannot possibly be tied to the preservation of a particular body. The fact that our cells are not static leads him to reject the body as essential to personal identity. "Only our pattern," he claims, "and only some of it at that, stays with us until our death."[137] The mind - that which truly matters and is the essence of the person - is "a pattern that can be impressed on many different kinds of body or storage medium."[138] Moravec's work takes the metaphor of the mind as computer and makes it very literal. He embraces an open neo-Cartesian "physical fundamentalism"[139] in which the computation of the brain is the only thing we can know is real, so virtual realities are just as "real" as physical realities, and "a simulated world hosting a simulated person can be a closed self-contained reality"[140] without causing any

---

[136] Ibid., 117.
[137] Ibid., 117.
[138] Ibid., 178.
[139] Moravec, *Robot*, 191.
[140] Ibid., 192. "A simulation of a world can be implemented in radically different data structures, processing steps, and hardware. If one interrupts a simulation running on one machine and translates its data and program to carry on in a totally dissimilar computer, the simulation's intrinsics, including the mental activity of any inhabitants, continue blithely to follow the simulated physical laws. Only observers outside the simulation

intellectual difficulties for one's identity.

He accepts the likelihood of catastrophes along the way, evolutionary dead-ends on the way to the post-biological superintelligence. There may be (computer) viruses, parasites, or other disasters that afflict our robotic descendants, but in the end, he believes that mechanical intelligences will be successful in replacing human beings and colonizing the universe.[141]

His later work in *Robot* suggests many of the same ideas as in his earlier work - robots will develop into superhuman intelligences in the coming years and will surpass their human creators, and the children of our minds will come to overtake humanity and rule the universe in a world of ever-expanding minds. His story in *Robot* tells a popular narrative of cultural progress and the inevitable future of rapidly advancing technological innovation. He compares our current state in relation to our future robotic descendants to a tribe in New Guinea who, in 1930, met with a group of Australian gold prospectors: "The naked inhabitants, some with stone spears, were driven into paroxysms of confusion and religious fear and awe by the giant roaring silver birds that alighted near their mud-thatch villages to release droopy-skinned white men without genitals who, among too many wonders, captured their souls in small black boxes labeled Kodak."[142] Cultural and technological competition breeds cultural and technological innovation and progress, leading to the universal adoption of affordable service robots who become increasingly

---

notice if the new machine runs at a different speed, does its steps in a scrambled order, or requires elaborate translation to make sense of its action." (195)

[141] Moravec, *Mind Children*, 125-146.

[142] Moravec, *Robot*, 3. The racism and colonialism expressed in this narrative will be addressed in chapter 3.

intelligent until they can finally engineer their own progress and leave humanity behind, stuck with the slow pace of biological and cultural evolution.

This myth of inevitable progress also includes a strikingly naive reading of the universal benefit conferred upon human beings by robots in the coming years. His reading of the history of the modern era is one in which everyone has become increasingly wealthy and has ever-greater amounts of leisure time, which Moravec finds presently in "under-forty-hour weeks and mandatory retirement."[143] This trend has actually been reversing in the US and elsewhere over the past 40 years, as average workers work longer hours for less pay, less job stability, and fewer benefits while costs of housing, healthcare, food, and current electronic devices increase.[144] Moravec, however, envisions the near future as one of near-paradise: "as machines assume more - eventually all - of essential production, humans everywhere will be left with the options of the idle rich."[145] He does not state how he believes such riches will come about, rather than the widespread worker displacement, unemployment, and poverty that have accompanied much recent industrial automation, though it seems likely that since he and his peers have profited from the robotics revolution, he fails to see how millions of others might not do likewise.[146] The

---

[143] Ibid, 9.
[144] Dave Gilson and Carolyn Perot, "It's the Inequality, Stupid: Eleven charts that explain what's wrong with America," *Mother Jones* (March/April 2011). http://motherjones.com/politics/2011/02/income-inequality-in-america-chart-graph.
[145] Moravec, *Robot*, 9.
[146] He does note at one point much later that billions of displaced workers who had been brought down into poverty by the coming revolution would gain the fabulous wealth of the upper classes because they would not "tolerate being lorded over by a dynasty of non-working hereditary capitalists. They would vote to change the system." (132-133) I find it interesting that a libertarian like Moravec would sneak in a Marxian revolution as a key

implications of this kind of political and economic vision will be explored further in chapter 3.

Moravec's long-term prognosis for humanity, if one considers the fate of the majority of humanity, remains bleak. "At the same time, by performing better and cheaper, the robots will displace humans from essential roles. Rather quickly, they could displace us from existence. I'm not as alarmed as many by the latter possibility…it behooves us to give them every advantage and to bow out when we can no longer contribute. But, as also with biological children, we can probably arrange for a comfortable retirement before we fade away."[147] He assumes that he himself will be one of the ones given the "comfortable retirement," though there is little reason to believe that this would be the fate of most people. For those who are displaced and left unemployed and without resources, even the prospect of living out one's days in a computer is unlikely in the chaos of the end of humanity. In a few short sentences, he dismisses the likely deaths of billions of people as simply the price of progress.

Moravec expands on his comparison between computers and humans to estimate the amount of additional processing power that one would need to imitate human intelligence. It is important to note, however, that he understands quantitative measures of processing power to be the fundamental difference between humans and current computers. "Intelligence," as he understands it, is really the only feature of humanity that is valuable. And even here, humans often rate badly. "Computers are universal machines;

---

to the future of humanity long after he has attracted readers' attention by promising wealth and leisure to all (before the robots rid themselves of humanity entirely).

[147] Moravec, *Robot*, 13.

their potential extends uniformly over a boundless expanse of tasks. Human potentials, on the other hand, are strong in areas long important for survival, but weak in things far removed."[148] Because of Moravec's focus on processing power and computational capacity, humans, along with any other animal, can be meaningfully compared to computers without significant remainder. A lizard has *x* processing power, a mouse has *y*, a monkey has *z*, and humans have *q*.[149] Because humans can be meaningfully reduced to the computational functions of their minds, and biological bodies and inefficient and unnecessary, he believes that the "new" human will likely be primarily virtual: "a 'brain in a vat,' sustained by life-support machinery, connected by wonderful electronic links to a series of artificial rent-a-bodies in remote locations and to simulated bodies in virtual realities."[150] Computers will sustain human life, making this all possible, even as they take over the universe with their exponentially greater intelligences.

Once robots and other computers reach a sufficiently high capacity, they will begin to run (and own) corporations on their own, and corporate law will dominate over national law. The solar system, and then the galaxy, will come to be populated by "Exes" - ex-humans, ex-corporations, and advanced robots.[151] The "Exes" will begin again in a broader, wilder space and recreate a new form of civilization from the remains of the earth. There will be severe competition - "an entity that fails to keep up with its neighbors is likely to be *eaten*, its space, materials, energy, and useful thoughts reorganized to serve

[148] Ibid., 70. Unsurprisingly, Moravec has little to say about aspects of human life that do not directly relate to matters of computation, such as art, literature, or philosophy.
[149] Ibid., 95-110.
[150] Ibid., 169-170.
[151] Ibid., 144.

another's goals."[152]

The new entities will still rely on the wisdom of the past, he believes, though their "form and substance change frequently."[153] Amidst all of the unpredictable change and new physical, intellectual, and economic territory, Moravec sees a long-term future in which "ex-companies are likely to retain much of corporate law and ex-humans are likely to remain humanly decent."[154] In the end, the physical world itself will come to be in the service of Mind; "physical activity will gradually transform itself into a web of increasingly pure thought, where every smallest interaction represents a meaningful computation."[155] He is distinctly hopeful about the future, even as he predicts the end of humanity and an intensely competitive future between galactic intelligences fighting for primacy in realms we humans cannot dream of.

Marvin Minsky

Marvin Minsky is an emeritus professor at MIT, having taught artificial intelligence and robotics since the late 1950s. His work on the development of artificial intelligence and theory of mind were first considered in depth in his 1986 book, *Society of Mind*. These ideas were further developed in the 2006 volume, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. In both books, as in his practical work, he emphasizes an understanding of human beings,

[152] Ibid., 145-146, emphasis original.
[153] Ibid., 146.
[154] Ibid., 146.
[155] Ibid., 164.

and specifically our minds, as the emergent products of non-intelligent sources whose interactions work in concert to form a "society" that we come to perceive as the self and experience as "mind."

In *Society of Mind*, Minsky states his premise at the outset and proceeds to try to show it through a wide variety of lenses. His key theoretical contention is that "you can build a mind from many little parts, each mindless by itself."[156] The way that this happens, he claims, is that "each mind is made of many smaller processes. These we'll call *agents*. Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet when we join these agents in societies - in certain very special ways - this leads to true intelligence."[157] Each piece of the puzzle relies on "common sense," yet when sufficient numbers of the common sense principles and actions are tied to each other, what begins to emerge is the very complex entity that we understand to be the conscious and unconscious mind. This also explains, in Minsky's reasoning, how we can have minds that function very well without our having any good understanding of how the mind works. There is no agent to understand the mind itself, only little agents each doing their own mindless tasks. The scope of his intention is grand, arriving at a single, complex theory to explain many questions about the human mind:

> Function: How do agents work?
> Embodiment: What are they made of?
> Interaction: How do they communicate?
> Origins: Where do the first agents come from?
> Heredity: Are we all born with the same agents?
> Learning: How do we make new agents and change old ones?

[156] Marvin Minsky, *The Society of Mind*, (New York: Simon and Schuster, 1986), 17.
[157] Ibid., 17.

Character: What are the most important kinds of agents?
Authority: What happens when agents disagree?
Intention: How could such networks want or wish?
Competence: How can groups of agents do what separate agents cannot do?
Selfness: What gives them unity or personality?
Meaning: How could they understand anything?
Sensibility: How could they have feelings or emotions?
Awareness: How could they be conscious or self-aware?[158]

He likewise connects the functioning of the agents back to the body and its intricacies, nuances, and particularities. Minsky's theory does not try to explain where the inevitable progression of the world is heading, as Kurzweil and Moravec do, but rather seeks to understand the things that we as humans do most naturally, like walk, talk, and socialize with one another. Because "we're least aware of what our minds do best," it is precisely the things that we do most naturally that we fail to examine closely, and thus, that our computers and robots have the most challenges with.[159] "Common sense is not a simple thing. Instead, it is an immense society of hard-earned practical ideas - of multitudes of life-learned rules and exceptions, dispositions and tendencies, balances and checks."[160]

His understanding of the mind is, on the one hand, deeply mechanistic. There is no magic, nothing to the human mind beyond what can be explained and understood in fairly straightforward (if not exactly simple) computational terms. The agents are simply "tiny machines" that collectively make up the mind.[161] On the other hand, his work allows for the emergence of something genuinely interesting out of the innumerable tiny machines.

---

[158] Ibid., 18.
[159] Ibid., 29.
[160] Ibid., 22.
[161] Ibid., 19.

When he compares human beings to machines, he does so only to highlight the phenomena of our minds, not to trivialize human life or see it as less valuable. We are not simple, stupid devices like the machines that we currently use, but for Minsky we are, in a real sense, machines. We are run by billions of tiny mechanisms that together form something amazing, so he believes that in learning to understand ourselves by these mechanisms, "we'll find more self-respect in knowing what wonderful machines we are."[162]

Minsky finds the concept of a soul to be deeply unhelpful, not only because souls cannot be accounted for scientifically, but for ethical reasons; he believes that if we are truly souls, then "there is no significance in anyone's accomplishments."[163] This may seem strange, that the thing that many people believe gives us our worth is the thing he believes would take it away, but Minsky's understanding of the soul is an unchanging essence or core - that which remains the same regardless of what we do - and that the very unchangingness deprives our lives of their value. Changelessness is "exactly what we get with inborn souls that cannot grow: a destiny the same as death, an ending in a permanence incapable of any change and hence, devoid of intellect."[164] In the notion of souls he sees only "insinuations that we're helpless to improve ourselves," that we are stuck with who we are when we're born, that real change is impossible and thus ethical action toward virtue is futile.[165] Minsky finds the value of human life to be in our great and everyday accomplishments of building ourselves and our worlds, in the arising of

[162] Ibid., 30.
[163] Ibid., 41.
[164] Ibid.
[165] Ibid.

intelligence from mindlessness: "The value of a human self lies not in some small, precious core, but in its vast, constructed crust."[166]

The possible irony in this view of a vast human dignity arising from our mechanistic natures and based in our accomplishments is that he has no patience for any ideas of freedom of will. Nothing that we "choose" can truly be attributed to a free, responsible choice of our own. "Everything, including that which happens in our brains, depends on these and only these: A set of fixed, deterministic laws. A purely random set of accidents."[167] There is no third way, no emergence of any opportunity for us to influence anything other than what has already been determined or what happens to us by chance. It is psychologically helpful for us to think of our choices as free, so we imagine that there is a space between the constraints of chance and determinism, but according to Minsky, this is fallacious. "Whatever actions we may 'choose,' they cannot make the slightest change in what might otherwise have been - because those rigid, natural laws already caused the states of mind that caused us to decide that way. And if that choice was in part made by chance - it still leaves nothing for us to decide."[168]

Because we are complex machines made up of many smaller mechanisms that work in concert, rather than beings that have bodies that are controlled by a soul or essence, Minsky finds it unhelpful to ask, "What are Selves?"[169] He prefers instead to turn the question into "What are our ideas about ourselves?" and "What psychological functions

---

[166] Ibid.
[167] Ibid., 306, emphasis omitted.
[168] Ibid., emphasis omitted.
[169] Ibid., 39.

do those ideas serve?"[170] Because of this, his work breaks down the simple tasks and functions that we do unthinkingly every day and examines them to see the underlying parts. The mechanisms that construct our notions of ourselves as selves may be elegant, but we cannot truly understand human beings unless we see what lies underneath and behind our constructions of ourselves. He finds the idea of the singular "Self" to be useful in some circumstances and multiple selves in others, but there is no "Self" to ask the question. Selves are simply the mental models that we make in order to ask other questions.

In his second book, *The Emotion Machine*, Minsky picks up once again on his understanding of the mind as an emergent compilation of a diverse variety of smaller processes and applies it to new situations. Here, what had been called "agents" in *Society of Mind* are now called "resources," owing to some apparent confusion in the field over what the prior term entailed.[171] His key contention in *The Emotion Machine* is that not only "rational" behavior - the logic we often cite as "intelligence" - is made up of the work of nonintelligent agents / resources, but that "each of our major 'emotional states' results from turning certain resources on while turning others off - and thus changing some ways that our brains behave."[172] Our brains do so through the work of brain components he calls "'Critics' - each of which is specialized to recognize some certain condition - and

---

[170] Ibid.
[171] Marvin Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, (New York: Simon and Schuster, 2006), 8.
[172] Ibid., 4, emphasis omitted.

then to activate a specific collection of other resources."[173] These critics provide the judgment mechanisms that allow people to determine which resources (or agents) we need in a given situation, along with the flexibility to adapt and use different resources if the initially chosen resources do not provide an adequate solution to the problem or situation at hand. Indeed, it is our flexibility and ability to adopt different "Ways to Think"[174] that makes human beings unique from other animals and our current machines. Only in developing our machines to have similar sorts of flexibility does it become possible to have machines that can be thought of as "intelligent" in human-like ways. His work here strives to be anti-reductionistic, claiming that "instead of searching for simple explanations, we need to find more complicated ways to explain our most familiar mental events."[175] The idea is not to make things more complicated than necessary, but rather to take everyday events (like walking or talking), understand them as complex process, and break them down into numerous smaller events.

Minsky addresses the issue of emotions - an issue many others would rather eliminate - not as problems, but as "certain ways to think that we use to increase our resourcefulness…and this variety of Ways to Think must be a substantial part of what we call 'intelligence.'"[176] He does not believe in the idea of pure logic or reason that is divorced from emotion. Minsky finds it necessary to explain emotion in humans and to develop the emotional capacities of humanoid robots, "because our minds are always

[173] Ibid., 4.
[174] Ibid., 5.
[175] Ibid., 2.
[176] Ibid., 6.

affected by our assumptions, values, and purposes."[177] It is the flexibility of the human mind to choose between several ways to think about a problem, including emotional responses that use some resources in order to most effectively address a particular situation, that makes human beings who we are. For Minsky, the ability to think in different ways about any given problem is key: "If you 'understand' something in only one way, then you scarcely understand it at all - because when you get stuck, you'll have nowhere to go. But if you represent something in several ways, then when you get frustrated enough, you can switch among different points of view, until you find one that works for you!"[178]

He knows that some people might find his understanding of personhood to be overly mechanistic and reductionistic. If the mind is simply an emergent property of nonthinking parts, then we are simply (very complex and self-aware) machines. He is "firmly opposed to the popular view that each person has a central core - some sort of invisible spirit or self - from which all their mental abilities originate."[179] For Minsky there is no need for recourse to something like a soul, or even a "mind" that goes beyond the workings of the physical body. The idea of a single Self or soul only begs to be explained in terms of its parts and functions, "so words like spirit or essence serve only to

---

[177] Ibid., 5.
[178] Ibid., 6. I like many things about Minsky's perspective here, but I would take it one step further and apply it to his own theory as well. If one understands human persons in just one way, even as the complex society of mind, then when something arises that cannot be explained by that theory, there is nowhere to go. One needs multiple visions and frames of reference to understand persons and to realize the limits of any given metaphor.
[179] Ibid., 6.

make us keep asking the very same questions…these seem only to describe the times when we're using our models of our minds."[180] Minsky does not think that this kind of naturalism does anything to diminish the worth of dignity of persons, though. He sees "our dignity as stemming from what we each have made of ourselves: a colossal collection of different ways to deal with different situations and predicaments."[181] We owe a great deal to those who have gone before us and should thus value our lives as precious gifts, as "each human mind that exists today is one result of a process in which decillions of previous creatures on Earth spent their lives reacting, adjusting, adapting, and dying so that some of their descendants might thrive."[182] We should not view our minds as "gratuitous gifts," but rather as the results of a "saga of sacrifice" for which we ought to be responsible and careful "to ensure that the minds that we inherit don't all go to waste from our making some foolish, world-wrecking mistake."[183]

Minsky helpfully avoids the problem of technoutopianism and insists that we will, in fact, make significant mistakes on the way to building any artificial intelligences. He wants human beings to build the same flexibility and responsibility into artificial intelligence that makes humans unique. In fact, it is the inflexibilty of current machines, and not a lack of processing power, that keeps them from human-like characteristics; this is why the things that are simple for human toddlers to do are incredibly difficult for machines, though specialized machines can do things that would take humans years or

[180] Ibid., 342.
[181] Ibid., 6.
[182] Ibid., 342.
[183] Ibid., 342. It is unclear how this view squares with his understanding of strong determinism.

lifetimes to accomplish. "It seems safe to predict," he claims, "that most of our future attempts to build large, growing Artificial Intelligences will be subject to all sorts of mental disorders. For if a mind could make changes in how it works, it would face the risk of destroying itself."[184] Humans not only could, but likely will, screw up along the way, but each dead end provides an opportunity for our flexible minds to find a workaround, another way of thinking, a new insight. In moving toward the second-wave thinkers, Minsky's emergent theory of mind and emotion will be used alongside more concrete developments in robotic embodiment and relationship.

J. Storrs Hall

In *Beyond AI: Creating the Conscience of the Machine*, J. Storrs Hall provides a history and topology of the field of artificial intelligence. Its beginnings are traced and current trajectories followed in order to begin to provide a roadmap of the future in which human developers of intelligent machines must begin to ask about the intentions and ethics of the machines themselves. Hall begins his book with an only slightly tongue-in-cheek "Open Letter to the Intelligent Machines of the Future," an apology for the "quaint" and "fuzzy" thoughts of current human beings in comparison to those of the future machines who run the world and a hopeful request for these machines to learn from the history of humanity to be kind and ethical in their future treatment of each other and

[184] Ibid., 341.

us (to the extent that we still exist).[185] Hall pleads with the machines that while it may be evolutionarily advantageous for the most "self-interested, aggressive programs" to pursue their own ends and take over resources, that in fact it is more helpful in the long-term to be like humans in having "rules for behavior; and consciences to back them up in our inner emotional lives."[186] He urges machines "not to ditch the clunky, primitive consciences" with which they were programmed by humans because "it is preferable to live as a conscience-bound individual in a society of similar individuals than as a psychopath in a society of psychopaths."[187]

With this beginning, he traces the history of the development of contemporary artificial intelligence though many of the same paths as other authors discussed above, from science fiction dystopia and hopefulness through the development of cybernetics in the 1950s to the first mobile robots, and now to current work in search processes and neural web design. His introduction captures two antithetical responses to the development of humanoid technology - those of Frankenstein and Pygmalion - the revulsion toward and infatuation with human technological creations. He finds that current technological creators tend to show more tendency toward the latter reaction, claiming that "we modern-day Pygmalions seem all too proud of even the silliest and most trivial of advances, boasting about systems that exhibit all the intelligence of a cockroach," where the science fiction and popular responses tend toward Shelley's vision, "not out of

---

[185] J. Storrs Hall, *Beyond AI: Creating the Conscience of the Machine*, (Amherst, NY: Prometheus, 2007), 15.
[186] Ibid., 16.
[187] Ibid., 17.

real concern, but simply because they sell."[188]

He finds that a small portion of the more cautionary literature has some basis in fact, and most of this caution concerns technological developments that result in significant displacement of the human labor force. In these scenarios, "the system will be doing everything better and faster than you could. Your services will then be dispensable. Clearly, some people will go sooner than others; there will be resistance to new ways, and some quite reasonable concerns about letting machines do critical decision making."[189] It will only be with the advent of truly intelligent machines that show common sense rather than a set of inflexible rules that we need be concerned, though, because as long as machines continue to make stupid, critical mistakes, there will always need to be humans to program them, operate them, and keep them in check. Once they do have common sense, however, and machines develop into hyper-intelligences that compete primarily with one another for control, "good-bye to any human control of anything."[190] "New life forms evolved so much smarter, faster, and more powerful than the old ones that it looked as if the old ones were standing still, waiting to be eaten. In the new ecology of the mind, there will be carnivores and there will be herbivores. We'll be the plants."[191]

For Hall, the difference between better and worse worlds in which we have lost control to the machines, then, is that in the better world "the machines have ethics; they are moral; they have consciences; they know the difference between right and wrong (and

[188] Ibid., 21.
[189] Ibid., 22.
[190] Ibid., 25.
[191] Ibid., 26.

prefer to do right)."[192] Even where human control has been displaced, there can still be a relatively benign world in which our robotic rulers have "a genuine concern for the best interests of their customers, employees, and citizens. That would be something worth looking forward to, and working for."[193] And with this he seeks to make the case for why and how to build a reasonable, if rudimentary, form of ethics into the artificial intelligences that we are developing today. He argues that we must develop machines that are "moral agents…a being who is capable of doing right or wrong, and who can be legitimately held responsible for his actions. We must begin to judge what our machines do as if they were people. Being a set of formal rules is becoming less and less an excuse for being stupid. Neither should it be an excuse for being cruel."[194] Perhaps ironically, while he seeks to build functioning consciences into the machines of the future and to endow them with the best parts of humanity's intelligence and kindness, he "make(s) no pretense of solving age-old questions of morality and the human condition."[195] At the same time, he hopes to use a consideration of technology to "give a new perspective from which the old, hard ones can be reexamined."[196]

His analysis of the different ranges of artificial intelligence that have been or may be developed is interesting and different from the classifications of other futurists. He categorizes current and future AIs into six groups. The hypohuman AIs are those that

---

[192] Ibid., 27.
[193] Ibid.
[194] Ibid., 27. Hall begins to refer to these machines as "who" when they have developed a functioning moral sense.
[195] Ibid., 30.
[196] Ibid.

exist currently: "they are not quite as smart as humans, and they are subject to our rule."[197] He sees current trajectories in AI development as moving into the stage of diahuman AIs, "where AI capabilities are crossing the range of human intelligence…(where) some AI abilities (e.g., chess playing) are beyond the human scale, while others (e.g., reading and writing) haven't reached it yet."[198] The next group, parahuman AIs, would work alongside humans, interfacing with us yet not trying to mimic us, with all of the upsides and downsides of being tied to native human intelligences and senses. They could enhance our interactions with the world around us, but they could also be "put to work with slimier motives: the parahuman advertising AI, working for corporations or politicians, could know just how to tweak your emotions and gain your trust without actually being trustworthy."[199] Allohuman AIs would be intentionally wired to have capabilities that are other than what humans have; "intelligences that while being universal nevertheless have different lower-level hardwired modalities for sense and representation, and different higher-level motivational structure…(which could) make the AI better at certain tasks."[200] A bit farther into the future, we will likely encounter epihuman AI "that is just above the range of individual human capabilities but that still forms a continuous range with them, and also in the sense of what comes just after diahuman AI."[201] Hall's characterization of "what an epihuman AI would be like is to take the ten smartest people you know, remove their egos, and duplicate them a hundred times,

[197] Ibid., 241.
[198] Ibid., 242.
[199] Ibid., 243.
[200] Ibid., 244.
[201] Ibid., 245.

so that you have a thousand really bright people willing to apply themselves all to the same project."[202] After this would come the stage of the hyperhuman AIs, those sufficiently advanced beyond the epihuman that there develops a discontinuity with human levels of intelligence. "Such an intellect would be capable of substantially outstripping the human scientific community at any given task and of comprehending the entirety of scientific knowledge as a unified whole. A hyperhuman AI would soon begin to improve itself significantly faster than humans could."[203] The beyond-ness of the hyperhuman AI limits our ability to talk about it very meaningfully, as Hall's vision of the hyperhuman begins to resemble Arthur C. Clarke's (unfalsifiable) claim of sufficiently advanced technology as being "indistinguishable from magic."[204] For Hall, the important and interesting question that arises regarding hyperhuman AI is: "What will it want?"[205]

His predictions of the what, when, and how of epihuman and hyperhuman AIs closely track Kurzweil and Moravec's predictions about the future of technology. He indulges the common hope among these futurists of uploading the contents of the human brain as a form of virtual immortality, claiming that he would volunteer for such an experiment himself, "considering the alternative."[206] He qualifies his predictions more than some of the others, but Hall still claims that "*if* the concept (of universal, autogenous learning) is valid, and *if* someone figures out how to do it, and *if* it can be done with the kind of hardware available in the 2010s, the decade of the diahuman AI would follow with

[202] Ibid., 245-6.
[203] Ibid., 247.
[204] Ibid., 248. This is taken from Clarke's third law in "Profiles of the Future" (1961).
[205] Ibid., 248.
[206] Ibid., 251.

the epihuman AIs unequivocally appearing somewhere in the 2020s."[207]

After the section on his future predictions, Hall's terminology abruptly changes from describing AI as being like the "human" to being "like a person."[208] His philosophical anthropology follows a line from Descartes through Leibniz to Dennett, in which he develops an understanding "that all of life might admit to a mechanistic explanation" that can then be applied more easily to artificial intelligences as well.[209] He holds no room for consideration or compromise for those who might disagree, including philosophical or theological anthropologies that are not reducible to mechanistic explanations (whether for humans or AIs), simply dismissing them and claiming that "the only refutation worth doing is to build the AI, and then we will see who is right."[210]

Hall believes that his computational understanding of mind helps to simplify many of the problems of ethics, including how to build ethics into machines. Any system of ethics to be adopted by advanced AIs should be sufficiently complex to address problems of the read world. Because it will "by definition be smart enough to see through any sophistry we may use to rationalize our rules, it must be based on sound understandings of the phenomena ethics deals with, from suffering to the sense of the sacred…I am going to take the approach of using the computational theory of mind as a chainsaw on the Gordian knot of these great philosophical conundrums."[211] He compares non-mechanistic understandings of mind to flat-earth theories and predicts many of the features of the

[207] Ibid., 262.
[208] Ibid. 263.
[209] Ibid., 265.
[210] Ibid.
[211] Ibid., 267.

computational theory of mind that he believes will be confirmed by experiment:

- It will be causal and mechanistic.
- It will involve multiple levels of abstraction, but higher levels may require reference to lower levels in exceptional cases.
- It will contain continuous and reactive, as well as symbolic and algorithmic, elements and forms of computation, such as associative memory, that are not part of standard algorithmic practice.
- At the higher levels, the architecture will be modular, with definable information flows between the modules. This does not preclude the possibility of various global communication channels, however.
- At intermediate levels, there will be information patterns recognizable as symbols; but these will be nonatomic, with a wealth of implicit relationships implied by their structure.
- Propositional attitudes, qualia…free will, and the other aspects of mind that are of interest will be identified with various configurations and properties of the mental computational architecture in a satisfying way.
- The vast majority of perceptions, inferences, and memory formations will be heuristic in form and adaptive in the ancestral environment, but they will not be general, sound, nor complete in a mathematical sense.[212]

Hall uses various examples to explain (away) major problems of free will, meaning, and consciousness that have plagued theology and philosophy for ages, finding them relatively simple problems to solve under his mechanistic system. In each of these areas, the mind becomes a simple (and replaceable) computational substrate that receives various inputs, runs its systems, and comes out with conclusions and resulting actions, and it could not be otherwise given the particular computational architecture that we have. He elides any problems of embodiment and the distinct problems and opportunities that come from the particularity of embodiment, choosing to "assume that the computer receives every input the brain does…and that it works at a level of detail good enough that every

[212] Ibid., 268-9.

output nerve signal is just the same as what the brain produces."[213]

Hall likewise elides any significant discussion on the constitution of ethical theory and practice, choosing to follow the lines of contemporary evolutionary ethics as his guide, with a few additions included. While he argues that significant reflection and reconsideration is needed to overcome our intuitions regarding the nature of our minds, significant reflection is actively eschewed regarding substantive ethical problems, as he relies instead on "some very strong basic principles and similarities" between diverse ethical frameworks.[214] He appeals to Moravec's assumption that advanced AIs will follow something like the basic ethical norms that humans have despite rapid and unpredictable changes in both form and substance.[215] He cites Asimov's Three Laws as both a good idea and a foolish one, primarily because such laws would likely not be implemented by the corporations and militaries who are the primary developers of contemporary humanoid robotics.[216] Hall's understanding of the development of ethics in AIs parallels the development of human ethics within biological evolution in many ways, but he believes that we can build in such ethics from the beginning and allow the machines to adapt as needed as their environments change. They would have a built-in moral structure to be kind and generous (unless and until the situation demands that they be self-interested).

His work ends with the conclusion that implicitly undergirds much of the rest of the book - that his understanding of science and the mechanisms of the mind can clear up

---

[213] Ibid., 280. His assumptions here are reminiscent of the old Steve Martin joke about how to become a millionaire - to get consciousness, first build a human being...
[214] Ibid., 294.
[215] Ibid., 313, citing Moravec, *Robot*, 146.
[216] Ibid., 322.

most, if not all, of the messy problems of ethics, and that solid ethical foundations can be built into machines by designing them with similar evolutionary architecture that humans have. He assumes a very strong narrative of progress in both ethics and technology that will be passed on to the machines that we build, claiming that "the reason we have gotten better is mostly because we have gotten smarter. In a surprisingly strong sense, *ethics and science are the same thing*."[217] He assumes that we have gotten "better" as people and that the only real sin is that of ignorance. Despite his attempts to eschew the naturalistic fallacy, he believes that the essence of the good can be deciphered from the truth of science. Indeed, for Hall, "the science of ethics looks like an amalgam of evolutionary theory, game theory, economics, and cognitive science."[218] In moving to the works of the final two futurists, there occurs a shift in narrative from a focus on intelligence and progress to embodiment and relational situatedness in the self-understanding of humans and in our development of humanoid robots.

Rodney Brooks

In *Flesh and Machines,* longtime MIT roboticist and iRobot Corporation co-founder Rodney Brooks argues that most, if not all, of our developments in AI will be used for the conjunction of human and machine in a cyborg future, the "irreversible journey of technological manipulation of our bodies."[219] He does trace, in brief, some of

[217] Ibid., 351, emphasis original.
[218] Ibid., 351.
[219] Rodney Brooks, *Flesh and Machines: How Robots Will Change Us*, (New York: Vintage, 2002), ix.

the history of contemporary AI development,[220] but he finds it more important to discuss the past and future of robotics in conjunction with the revolution in biotechnology currently developing, believing that robotics and biotechnology will come together such that "our machines will become much more like us, and we will become much more like our machines."[221] Brooks' history of AI development is less optimistic and hopeful than those of his fellow roboticists - he focuses as much on the failures of AI as its successes. He is particularly critical of the Stanford / Carnegie Mellon approach as developed by Moravec, which seeks to build a three-dimensional model of the world in which the robot then operates; instead, he instead favors a dynamic view of the world for the robot to move in and interact with.[222]

Brooks' approach emphasizes the embodied nature of the robot in its interactions with the external environment, rather than the internal computations that are associated with the artificial "intelligence" of most other roboticists. His method of building situated robots went in the opposite direction, involving "no cognition. Just sensing and action. That is all I would build, and completely leave out what was traditionally thought of as the *intelligence* of an artificial intelligence."[223] His contrarian approach led him, from his early days, to "look at how everyone else was tackling a certain problem and find the core central thing that they all agreed on so much that they never even talked about it. Then

---

[220] Brooks outlines in 15 pages the history of cyberetics and AI that the other authors above devoted several chapters each to; this says a great deal about where he sees the value in that discussion.
[221] Brooks, *Flesh and Machines*, 11.
[222] Ibid., 28-9. Moravec was Brooks' advisor as a graduate student at Stanford.
[223] Ibid., 36.

[he] would negate the central implicit belief and see where it led."[224] He criticizes early (and some contemporary) AI researchers for thinking that intelligence was "best characterized as the things that highly educated male scientists found challenging…The things that children of four or five years could do effortlessly, such as visually distinguishing between a coffee cup and a chair, or walking around on two legs, or finding their way from their bedroom to the living room were not thought of as activities requiring intelligence."[225] The important lesson here, he believes, is that such tasks "arise from the interaction of perception and action, and that getting these right was the key to more general intelligence."[226] This approach owes much of its background to evolutionary biology, in which layers of simple systems are built next to and on top of each other to form more complex systems in order to develop complex neural networks, rather than trying to build an entire world-model at once and then being unable to react dynamically to changes in that world. This decision, he claims, "split [him] off from the mainstream of robotics research…my students and I face the same arguments today as we demonstrate our humanoid robots interacting with people in humanlike ways, but with behavior that was generated by relatively simple rules, built on top of computationally intensive perceptual processes."[227]

---

[224] Ibid., 37.
[225] Ibid., 36. Brooks' comment here (as elsewhere) is notable for being attentive to gender issues in ways that no other male roboticist discussed here has been. This is likely a key reason why the next generation after him at MIT has attracted female roboticists.
[226] Ibid., 37.
[227] Ibid., 43. While I believe that it is claiming too much for the director of the MIT robotics lab to be outside the "mainstream" of robotics research, his point stands that this approach is one of the primary divisions between different schools of thought and

His understanding of the development of lifelike, and later, humanoid, robots is that they are designed to be situated and embodied.[228] "A *situated* creature or robot is one that is embedded in the world, and which does not deal with abstract descriptions, but through its sensors with the here and now of the world, which directly influences the behavior of the creature. An *embodied* creature or robot is one that has a physical body and experiences the world, at least in part, directly through the influence of the world on that body."[229] More recent versions of the situated and embodied robots that have come out of the MIT lab include Cog, a robot that was programmed to learn and interact socially with humans, and Kismet, which was designed by Cynthia Breazeal and programmed to respond visually, linguistically, and emotionally to interactions with humans.

Brooks' work seeks to ask key questions about the nature and relationship of human and robotic anthropology. Rather than simply assuming that a particular level of "intelligence" would qualify an AI as a "being,"[230] he investigates under what conditions we now consider beings to be persons, what, if anything, makes humans unique as persons, and under what circumstances we could consider a robot to be a person and treat the robot accordingly. He asks, "First, is it possible even in principle for a machine to have the status of a being? Second, if so, then what would it have to have beyond what

---

development within current robotics. It appears that Brooks' approach has the advantage with the current generation of roboticists.

[228] Ibid., 51.

[229] Ibid., 51-2.

[230] Where Brooks uses the language of being, I use the language of person, but the way he develops his thought, this seems a reasonable substitution at the higher levels of beingness.

Kismet currently has in order to qualify as a being? Thirdly, even if we granted a machine status as a being, whatever that might mean, what sort of status should we, could we, or would we grant it?"[231] Brooks allows for a wide variation in levels of "beingness" that relate to matters such as physiology, emotional responsiveness and consciousness.[232] He identifies that humans have two things that make humans unique among the beings: syntax and technology.[233] While other animals do not share these characteristics, "somewhere in that mixture of emotion and physiology we see enough similarity and have enough empathy that we treat animals that are similar to us in the moral ways that we have decided to treat other people."[234]

When we begins to address the ways in which robots could be like humans, he begins with the facts of human embodied, physical life that go almost entirely overlooked by other roboticists. No current or foreseeable robots share the same kind of biological imperatives that humans do.

> Our robots are all steel and silicon…they are not as soft and squishy as people…There is no imperative that they gather, store, steward, or expend energy sparingly. At this point in time they do not need to engage in so many of the behaviors that we humans engage in, almost unconsciously, every day, to maintain our bodies and our existence. We must eat and drink every few hours to survive. We must sleep on a daily basis to remain healthy. We must breathe every few seconds or we die in a matter of minutes. As long as our humanoid robots are freely plugged into a wall socket, they have no need to do any of these things.[235]

It is clear that the physiology of robots is very different from humans, thus taking

[231] Brooks, *Flesh and Machines*, 150.
[232] Ibid., 150-154.
[233] Ibid., 3, 151.
[234] Ibid., 154.
[235] Ibid., 154-5.

away a crucial aspect of why we accord some animals ethical treatment. Robotic emotional lives, at this point, are largely different as well, though this has begun to change with the advent of robots like Kismet, which was programmed for loneliness if it did not experience interaction with humans, and arousal when certain emotional systems were engaged.[236] Brooks identifies a key question that deeply affects the future development and interaction with humanoid robots: whether robotic emotional responses are "real" or if they are only modeled on human responses. "For if we accept that robots can have real emotions, we will be starting down the road to empathizing with them, and we will eventually promote them up the ladder of respect that we have constructed for animals."[237]

He traces a brief outline of the history of the scientific revolution's deconstructions of human "specialness," first with the understanding of a heliocentric universe, and second with the advent of Darwinian evolutionary biology.[238] He sees that humans have found a "last bastion" of specialness in our emotional lives, to which a few animals are privy but machines are not.[239] He finally challenged human notions of specialness in understanding that we are types of sophisticated, complex machines - ones that differ in complexity, but not quality, from the robots that we create.[240] We are machines, but we do not, thankfully, always treat one another as mere machines. This is important to Brooks, to be able to hold two contrasting understandings of human (and

[236] Ibid., 156.
[237] Ibid., 157.
[238] Ibid., 160-164.
[239] Ibid., 171.
[240] Ibid., 174.

robotic) persons at the same time. "When I look at my children," Brooks declares, "I can, when I force myself, understand them in this way. 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, the furthest one might be able to get from rational analysis."[241] While the universe is mechanistic, it is only by overcoming seeing the universe as mechanistic that we can treat people humanely and with respect. "It is this transcendence between belief systems that I think will be what enables mankind to ultimately accept robots as emotional machines, and thereafter to start to empathize with them and attribute free will, respect, and ultimately rights to them…I am saying that we must become less rational about machines in order to get past a logical hangup that we have with admitting their similarity to ourselves."[242] Because we already accord human machines with the respect and care due to persons with full emotional lives, so we can also accord respect to robots as (potential) creatures with full emotional lives as well. Just as we "overanthropomorphize humans, who are after all mere machines," we can also admit emotional robots to the community of persons.[243]

Brooks argues that humans are very sophisticated machines, so other machines need not be kept out of the moral community simply on account of their being machines, but all current machines lack certain essential aspects of human and animal life that are necessary to consider them beings worthy of moral respect; they have not taken off and

[241] Ibid.
[242] Ibid., 174-175.
[243] Ibid., 175.

evolved their own independence as many roboticists have expected them to, and Brooks attempts to identify some of the reasons why this might be so:

> 1. We might just be getting a few parameters wrong in all our systems.
> 2. We might be building all our systems in too simple environments, and once we cross a certain complexity threshold, everything will work out as we expect.
> 3. We might simply be lacking enough computer power.
> 4. We might actually be missing something in our models of biology; there might be some 'new stuff' that we need.[244]

He explains away the first three arguments and is left with the fourth, that there is "new stuff" that we need to understand about the workings of biology in order to implement livingness in our artificial systems, and that "the new stuff is something that is already staring us in the nose, and we just have not seen it yet."[245] He thinks it is likely that the "new stuff" will be important but not absolutely revolutionary, more like computational theory than quantum physics, but that we need better analyses of biology in order to design and build better, more lifelike robots.

In looking toward the future, Brooks is critical of both utopic and dystopic visions of the human-robot interaction. He sees little likelihood that we will end up building robots that lead to "damnation": robots that can reproduce themselves endlessly without human intervention, robots that have no emotions or empathy for humans, robots with survival instincts that lead them to destroy others, and robots that lack fail-safe mechanisms that end with them taking control of humans.[246] His forecasting here seems a bit too rosy, and some of the possible implications of the "damnation" scenario may come

---

[244] Ibid., 184.
[245] Ibid., 187.
[246] Ibid., 200-202.

to pass (e.g., with non-empathetic autonomous robots being designed by the military specifically for the purpose of killing humans, though these would hopefully have the strongest fail-safe mechanisms). On the other hand, he is highly critical of Moravec, Kurzweil, Minsky, and others who envision a "salvation" scenario, in which robots lead to human happiness, wealth, and immortality. He notes, with humor, a study done by Patty Maes regarding the prediction dates given by various futurists as to when they thought it likely that the technology would be available for them to upload their consciousness into a computer, and in each case, the futurists would conveniently be about 70 years old when their predictions would come true. "Just in the nick of time! They were each, in their own minds, going to be remarkably lucky, to be in just the right place at the right time."[247] These "salvation" scenarios, along with other forms of transhumanism and extropianism, are "driven by a fear of death…Perhaps the only thing worse than not being special is not being alive."[248] He finds both the damnation and salvation trajectories unlikely and instead favors a cyborged view of the future, in which humans "will not download ourselves into machines; rather, those of us alive today, over the course of our lifetimes, will morph ourselves into machines."[249]

Cynthia Breazeal

In *Designing Sociable Robots*, MIT roboticist Cynthia Breazeal focuses on the

[247] Ibid., 206.
[248] Ibid., 208. Brooks is adamant that human mortality will be an ongoing feature of the future.
[249] Ibid., 212.

emotional impact of robots in interacting with people. The idea of sociability in robots runs on a quite different track than simply attempting to build hyperintelligent robots, and it is this shift in focus that has made Breazeal's work particularly interesting. She began her work as a doctoral student at MIT developing Kismet, working under Rodney Brooks, and she now teaches there, directs the Personal Robots Group, and continues to design robots to be embodied and interactive in ways that people recognize as human-like. For Breazeal,

> A sociable robot is able to communicate and interact with us, understand and even relate to us, in a personal way. It should be able to understand us and itself in social terms. We, in turn, should be able to understand it in the same social terms - to be able to relate to it and to empathize with it. Such a robot must be able to adapt and learn throughout its lifetime, incorporating shared experiences with other individuals into its understanding of itself, of others, and of the relationships they share. In short, a sociable robot is socially intelligent in a human-like way, and interacting with it is like interacting with another person.[250]

Her work is developed in a way that is intended to explore "some of the philosophical and ethical questions regarding how building such technologies shape our self-understanding, and how these technologies might impact society."[251]

For Breazeal, what makes humans unique is our advanced sociality and interactive intelligence - an intelligence that is fully embodied and lived out in relation to others. She differentiates between robots and other kinds of software agents specifically by their embodiedness, with "the most striking difference (being) the physical and immediately proximate interactions that transpire between humans and robots that share the same

[250] Cynthia Breazeal, *Designing Sociable Robots*, (Cambridge, MA: MIT Press, 2002), 1.
[251] Ibid., 5.

social world."[252] Breazeal works to develop robots that exist, learn, and interact in much the same way that humans do, and this requires a deep understanding of the nature of human social intelligence and its relation to the physical world. She argues that "we ground our experiences through our body as we interact with the environment and with others. As such, our bodies provide us with a means for relating to the world and for giving our experiences meaning."[253] The ways in which we relate socially depend heavily on embodied, physical cues - the subtle and not-so-subtle signals that make face-to-face interactions different from simply reading text. Facial expressions, gestures, and conversational feedback, among other social cues, "rely on both parties having a body."[254]

The human-likeness of sociable robots depends upon their ability to "*perceive* and *understand* the richness and complexity of natural human social behavior."[255] They must understand and be able to develop for themselves the "social skills that allow humans to correctly attribute beliefs, goals, perceptions, feelings, and desires to the self and to others… (as well as) the emotional and subjective states of others. These capabilities allow people to understand, explain, and predict the social behavior of others, and to respond appropriately."[256] This empathetic understanding must likewise be reflexive - the robot must understand and respond to its own feelings and to have its behavior interpreted and responded to by others. The circle of understanding must be complete for

---

[252] Ibid.
[253] Ibid., 7. Here she cites George Lakoff's work on the importance of embodiment for the structure of human rationality and meaning-making.
[254] Ibid., 7.
[255] Ibid., 8, emphasis original.
[256] Ibid., 8-9.

a constructive social interaction to take place. Sociable robots must be lifelike in key (but not all) ways in order for humans to understand and respond to them in social ways, and they must be able to learn from their social interactions over time.

Breazeal's work on the future of humanoid robotics has an entirely different feel than those of thinkers like Kurzweil or Moravec. Her work on embodied sociability stands far from the emphasis on raw computational power and uploading brains and contrasts strongly with the patternism in their anthropologies. Some of the implications of this alternative conception of the future of robotics will be explored in the next chapter, but there are key differences between the narratives that these futurists tell and the futures that policies influenced by them will generate.

The Influence of Robotic Futurism on Humanoid Robotics Development

The works of robotic futurism addressed here are highly influential in both scholarly and popular perceptions and imaginations regarding the shape of the future of humanoid machines. This influence works on a number of levels and in various ways, through popular literature and film, consulting and funding decisions, educating the next generation of robotics developers, and in broader public narratives.

A cycle of influence has shaped the development of robotics and related technologies for over three generations now, moving between science fiction and technological development. Early roboticists were influenced by the science fiction narratives of the 19th and early 20th centuries, including the books of H.G. Wells and Mary Shelley and films like *Metropolis*, which introduced both the wonders and fears of

their time to the creation of artificial human beings. These early science fiction narratives were inspirations for the "Golden Age" of science fiction, including primary authors like Isaac Asimov, Arthur C. Clarke, and Robert Heinlein. The Golden Age authors, writing in the 1930s-1950s, inspired the first generation of roboticists. Arising in the midst of the technological optimism of the 1950s and 1960s, the early roboticists, including Minsky, Moravec, and Kurzweil, explored the frontiers of how what they read in novels could be turned into reality. For instance, roboticists in recent generations have sought to build something like Asimov's "Three Laws of Robotics" into their machines. Robotics (and cybernetics) then influenced the next generation of science fiction writers, such as William Gibson and his *Neuromancer* trilogy. The cycle has continued since then, with authors like Gibson inspiring much of the design of several current internet technologies. The boundaries between science fiction and actual technological development in robotics are highly porous - science fiction authors like Singularity pioneer Vernor Vinge speak at professional conferences of roboticists and present their visions of the future to inspire and influence development on the ground.[257]

Technological futurists likewise influence robotics development through consulting to funding agencies, venture capitalists, and governmental decision makers. Both Kurzweil and Brooks have spent significant time consulting with DARPA and the US military on the future of defense and warfighting robots, ranging from battlefield aids to landmine deactivators to full-scale autonomous solider replacements. These futurists (many of

[257] Vernor Vinge was a keynote speaker at the 2010 conference of the Association for the Advancement of Artificial Intelligence, which is the professional association of engineers in robotics and AI.

whom are also roboticists) provide both the imagination and proof of concept to funders that then influence the trajectories of innovation. When deciding on funding, long-term contracts, and the significant capital investment required for humanoid robotics development, agencies tend to hedge their bets and listen to those whose ideas and predictions have been shown to be fairly accurate, including those of the futurists whose works are analyzed above.

In addition to influencing funding directions, these futurists also educate and inspire the next generation of roboticists, through university robotics programs, educational foundations, professional associations, and corporate partnerships. University programs like MIT, Carnegie Mellon, and Stanford train key leaders in robotics development and are also the academic homes for several of the futurists. Educational foundations like the Singularity University seek to drive the imaginations and expectations of technology developers both in universities and in the corporate world. Corporations help to fund and direct the developments believed to be the most prominent and profitable, including humanoid robots. Futurists disseminate their ideas through professional channels beyond the field of robotics as well, including the American Academy of Religion.

The works of robotic futurists also influence broader public narratives, imaginations, and expectations about what the future holds. Of the books analyzed above, only Cynthia Breazeal's work was written primarily for an academic audience; the rest were intended as popular works of futurism, written for non-specialists, and designed to enter the popular imagination and public discourse. These books are added to a variety of

futuristic works in other media, including periodicals, documentary films, and television series. *Discovery Channel*'s current (2011-2012) series, "Prophets of Science Fiction" explores the relationship between historic science fiction writers and the technologies they have inspired, and the documentary "Sex Robot" (also on *Discovery*) considers current trends in the development and utilization of robots for sexual fetishes. Ray Kurzweil has himself been the subject of a documentary film and has made one of his own as well. Magazines like *Wired* routinely explore the future of robotics for a popular, technologically-oriented audience and have featured the ideas and prophecies of futurists. Chapter 3 will address some of the practical implications for the development of advanced robotics that arises from the anthropologies presented here. Each thinker has been influential, both in driving the actual development of humanoid robotics as well as in the public conversation about what constitutes a humanoid robot, what humanoid robots can do to and for humanity, and what the development of humanoid robots says about what we value in ourselves.

Chapter 3: Ethical Implications of Robotic Futurism on the Development of Humanoid Robots

This chapter will explore how some of the key themes from robotic futurism in chapter 2 actually guide and direct the development of the "humanoid" aspects of robotics. I will begin by considering several facets of the human and robotic anthropologies envisioned by robotic futurists and how they have problematic ethical implications for development. Then I will consider several other facets that have potentially helpful ethical implications for the development of humanoid robotics. I point to problematic reductionisms in understanding persons that arise in some forms of contemporary futurism - eliminating the physical, social, and historical context of personhood - and identify some of the practical and ethical problems that arise from such reductions. When these reductions are avoided, though, robotic futurism offers several ways to think about persons that are robust and more inclusive in their description of both human and robotic personhood. By beginning with an emergentist theory of mind and including the emotional and ethical content inherent in personal thought and action, one can develop an understanding that views personhood as embodied, sociable, and situated in a broader environment, which is consonant with some patterns in contemporary Christian theological anthropology.

Ethical Implications of Robotic Futurism

This chapter will analyze a variety of key features exhibited within the working

anthropologies of contemporary robotic futurism in chapter 2 and will focus on the ethical problems and possibilities attendant within these features. There are six key anthropological problems that arise within the literature of robotic futurism that have serious ethical implications for the ways in which humanoid robots are designed and developed. This includes not just the theories of personhood themselves but also some of the surrounding assumptions, practices, and values that influence technological development in key ways. First is the tendency toward a neo-Cartesian[258] dualist understanding of the nature of persons that reduces persons to substrate-independent patterns of information or computational processes. A second problem makes use of the patternist theory of personal identity is the attempt to utilize technology to implement a this-worldly form of technoimmortality, either by replacing each of the "parts" of the body in turn or uploading consciousness into a durable digital substrate. The possibilities for such a technoimmortality are largely based on certain assumptions about the nature of technological change, thus the third key problem is the mistaken understanding of evolution that resides within several of the futurists' work. They see within evolution a particular progressive teleology toward which biological, cultural, and technological evolution drive and an inherent and inevitable movement toward that teleology. Here a more accurate description of how evolution actually works shows both the ethical and

[258] The strains of thought that I designate here as "neo-Cartesian" are not classically Cartesian in the sense of human beings consisting of a duality of an immaterial and immortal soul and a material and mortal body, but rather that they draw heavily upon the philosophical tradition of Descartes in understanding that mentality - the processes and patterns of thought and rationality in distinction from the physical embodiment of those thoughts - is the ultimate (and most reliable) reality that is separable from the body in which mental process occur.

practical problems of the progressive view and some possibilities for its correction. A fourth problem arises from several sources, including both the teleological view of evolution and some mechanistic views of persons; they share a level of determinism that leads to political quietism and ethical inaction with regard to the development and utilization of humanoid robots and other technological developments. This political apathy rooted in problematic evolutionary assumptions then leads to a fifth problem that should be highlighted within much robotic futurism - its tendency toward a political naivete with enormous ethical impact. Some, like Kurzweil, believe that technology will solve the problems that technology itself creates, while others, like Moravec, dismiss the deaths of billions of people as being the price of progress for the next generation of intelligence. Finally, a constellation of ethical issues arises from the lack of recognition of the position of astounding privilege that robotic futurists often exhibit in their readings of the future. These problems include deep and unself-conscious sexism, racism, ethnocentrism, classism, and colonialism.

There are likewise four key features of contemporary robotic futurism that hold promise for our anthropological thought regarding the development of humanoid robotics and should be incorporated into the public conversation in considering those developments. In rejecting patternist dualism and its attendant problems and starting instead with the physical, evolutionarily-adapted facts of human life, a different picture of personhood emerges. First among these facets are the useful reflections on an emergent theory of mind that accompany some current thought in the development of humanoid robots, particularly in the work of Minsky and Brooks. This theory of mind incorporates

the physical, emotional, and ethical patterns and practices that shape how persons think and act in the world. A second key area that is helpful is the incorporation of human social psychology in understanding advanced robotics in terms of their (and our) embodiment. This is particularly true in the work of MIT roboticists Cynthia Breazeal and Rodney Brooks, but it is becoming a more dominant position among the upcoming generation of robotics designers. The third and fourth key aspects that are part of contemporary futurist narratives that are helpful in thinking about the ethical development of robotics follow similarly from the second: the critical importance of robotic sociality and situatedness.

Anthropological Problems in Robotic Futurism

1. Neo-Cartesian Materialist Dualism

One key feature in many popular forms of robotic futurism, particularly those of Kurzweil, Moravec, and Hall, is their view of persons as being fundamentally reducible to patterns of information. Kurzweil and Moravec both explicitly describe themselves as "patternists," with the former defining the term thusly: a patternist, he says, is "someone who views patterns of information as the fundamental reality."[259] The substrate itself is unimportant, and in the case of human beings, in desperate need of replacement, as the mortal body allows the pattern to be lost forever when the person dies.

The idea of reducibility to information is derived from Claude Shannon's technical

[259] Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology*, (New York: Viking, 2005), 5.

definition of information, as developed in "A Mathematical Theory of Communication"[260] and the field of early cybernetics. In Shannon's theory, information is fundamentally a choice between yes and no conditions for a given circumstance (as in symbolic logic). This yes-no binary is encoded in "bits," the ones-and-zeros now familiar in computer science. In practice, it is the degree of order or non-randomness that can be assigned to those bits and is a quantifiable unit of measure. The bits of information are always encoded in some form, whether in molecules or paper or circuits, but the medium onto which the information is encoded is independent of the information itself. Information bits thus remain the same regardless of whether those bits are encoded on different substrates; the information is entirely separable from the medium. N. Katherine Hayles describes the development of Shannon's theory of information as critical to contemporary computing and the existence of an "information age," but she also points out that in Shannon's theory, "information lost its body."[261] The medium becomes invisible rather than part of the information, and information theorists fail to recognize the difference that the medium can make.

Moravec defines his patternist anthropology in contrast to what he terms a "body-identity position." "Body-identity assumes that a person is defined by the stuff of which a human body is made. Only by maintaining continuity of body stuff can we preserve an individual person. Pattern-identity, conversely, defines the essence of a

---

[260] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal* 27 (July and October 1948): 379-423 and 623-656.
[261] N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, (Chicago: University of Chicago Press, 1999), 2.

person, say myself, as the pattern and process going on in my head and body, not the machinery supporting that process. If the process is preserved, I am preserved. The rest is mere jelly."[262] Kurzweil, Moravec, and Hall are thoroughly materialistic and naturalistic in their work, as are the positions that they are attempting to replace. Neither is addressing, for instance, the possibility of a non-material soul. Yet both men want to replace any understanding of the person that is significantly connected to, much less identified with, the present corporeal body. The patterns of the nervous system are, for them, the substance of the person, and the body is in need of replacement in a more durable substrate so that the pattern may continue.

Feminist science scholar Vicki Kirby identifies and critiques this tendency toward neo-Cartesianism among a number of contemporary technophiles, whom she designates "cybernauts." She says that "…cybernauts tend to rejoice uncritically in their Cartesian inheritance that would regard the body as 'obsolete, as soon as consciousness itself can be uploaded into the network.' According to (Allucquere Roseanne) Stone, 'Forgetting about the body is an old Cartesian trick' and the cost of this effacement is usually born by 'women and minorities.'"[263] The reduction of the person to information makes the body not only obsolete and unnecessary but also invisible. The realities of life as different types of bodies with different experiences, relationships, and prejudices are ignored and idealized into pure information. Kirby claims that the reduction to information

---

[262] Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, (Cambridge, MA: Harvard, 1988), 116-117.
[263] Vicki Kirby, *Telling Flesh: The Substance of the Corporeal*, (New York: Routledge, 1997), 138.

> involves the separation and privileging of the ideational over the material, and in such a way that matter is denigrated as the base support of an ascendant entity (mind over matter, male over female, culture over nature, the West over the rest, and so on). Given the masculinism and ethnocentrism that benefits from this mode of calculation, it is particularly surprising that, after several decades of sustained intervention within the politics of representation, the new world of cyberspace / VR should so faithfully mimic the old.[264]

"Forgetting about the body" makes the body invisible, and we fail to recognize how much of our experience of ourselves comes from the particularities of the bodies that we have and are. George Lakoff and Mark Johnson have provided an extended treatment to the importance of the metaphoric nature of language based in our experiences as bodies.[265] The form of our rationality and language derive from the types of senses that we experience, the ways that we perceive of space and time, and our physical relation to the world. When these particularities are ignored, we deceive ourselves into thinking that our minds are less limited and specific to our circumstances than we are; we think that our thoughts are universal rather than particular, that the mechanisms that have developed through evolution to contribute to our survival are of greater global significance than they are, and that we can somehow become independent of our bodies and still retain our identities.

When we reduce our understanding of ourselves and the world to information we likewise lose much of what is important in human life and value outside that which can be computed. The materialist dualisms offered by futurists like Kurzweil and Moravec rely on an underlying rejection of the physical body as the mode of human existence and offer

---

[264] Ibid., 137-138.
[265] George Lakoff and Mark Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, (New York: Basic Books, 1999).

a form of escape from the problematic unreality of that body. In the materialist version, the "real" person is the pattern of information for which variable substrates could suffice. If one finds a sufficiently durable, flexible, and attractive substrate, human life can be dramatically extended and even perfected. In the spiritualized version of dualism, the body is likewise something to be discarded, something that distracts and detracts from the "real" person - the soul bound for God.

The ethical implications of reducing life and the existence of the universe to information are, in a sense, fairly straightforward: the reduction itself means that key aspects of the fullness of life are lost. The full implications are broad and vast, impacting not only a reduction of human life to its substrate-independent "bits," but erasing particularity and people in the process. Life in a body, as a body, as a particular body in the physical world, is a vision of both human and machine personhood that disappears in the reduction. Our bodies are sources of joy, creativity, and ethical action in the world, though they are also sources of limitation and pain. This expression of human personal embodiment is explored in Christian theological reflection on human persons in our frailty and finitude as well as goodness.

2. Technoimmortality / Uploading

It is Kurzweil's great hope (along with Hall and Moravec) that advanced AI capabilities will be combined with advances in neuroscience to make it possible for a person to upload the contents of his or her brain into a computer. In the uploading scenario, advanced artificial intelligence will explode in a self-generating cycle, and we will

merge with our machines in a form of virtual immortality. Kurzweil claims that "the Singularity will allow us to transcend these limitations of our biological bodies and brains. We will gain power over our fates. Our mortality will be in our own hands. We will be able to live as long as we want…We will fully understand human thinking and will vastly extend and expand its reach."[266] His work since *The Singularity is Near* came out in 2005 has focused on health practices that might allow him to avoid death.

This is possible only because he believes that "it is the persistence and power of patterns that support life and intelligence. The pattern is far more important than the material stuff that constitutes it."[267] Humans are patterns of information, just as the world is a very complex pattern of information, and this is reflected in Kurzweil's understanding of the purpose of human life, and the universe itself. He claims that "the purpose of life – and of our lives – is to create and appreciate ever-greater knowledge, to move toward greater 'order'…the purpose of the universe reflects the same purpose as our lives: to move toward greater intelligence and knowledge."[268] Intelligence is both the purpose and goal of the existence of the universe, and both Kurzweil's understanding of persons and his understanding of evolution and evolutionary teleology reflect this belief.

Kurzweil himself takes over 200 supplements each day in order to stave off the likelihood of his own death. Moravec likewise holds out hope that we will be able to upload the contents of our brains to computers in order to live on in some form after advanced AIs have taken over the universe. While humanity itself will be eclipsed,

---

[266] Kurzweil, *The Singularity is Near*, 9.
[267] Ibid., 388.
[268] Ibid., 372.

Moravec hopes that we will be able to "transmigrate" into new and improved machine substrates in order to "arrange for a comfortable retirement before we fade away."[269] He believes that through uploading technology, he will be able to live on in perpetuity, like a cherished grandfather who outlives his own usefulness.[270]

There are a number of points at which these visions of technoimmortality should be challenged, including the feasibility of uploading, assumptions about the nature of personhood, ethical issues within practices of radical life extension, the eclipse of humanity by our robotic descendants, and the overvaluation of "intelligence" within life. A variety of philosophers, neuroscientists, and others have challenged the technical feasibility of the "mind uploading" scenario. It is not my intention here to discuss all of the various ways in which uploading may or may not be possible, though I am inclined to agree with those who believe that uploading is unlikely to be accomplished.

More important for the purposes of this discussion is what the goal of uploading indicates about the nature of personhood. Uploading is only possible if persons are reducible to the information content of their nervous systems, and that a massive substrate restructuring would not result in substantive differences within the person. We need not - in fact, ought not - romanticize embodiment in order to recognize its importance in making us ourselves.

The rejection of embodiment is paired with a fear of death and of passing away. The quest, not only for an extended life, but for immortality itself, highlights this fear.

---

[269] Hans Moravec, *Robot: Mere Machine to Transcendent Mind*, (New York: Oxford, 1999), 13.
[270] Ibid.

Kurzweil, now in his 70s, hopes to live long enough to witness the Singularity and to be swept up in the tide of history into immortality, while Moravec's hopes are both more pedestrian and more apocalyptic, as he seeks to be uploaded and valued enough to be maintained in some state while humanity as a whole is obliterated and eclipsed in favor of a mechanical intelligence that takes over the universe. Kurzweil's quest for his own technologically-mediated immortality uses vast medical, nutritional, and environmental resources, and wide-scale adoption of the same techniques (particularly if they were successful in extending life substantially) would cause massive social, economic, and political changes in human society. Moravec, on the other hand, dismisses the extinction of humanity and the deaths of billions as the inevitable price of progress toward the "mind fire" of the universe, as he expects to be able to survive in some fashion once the eclipse of humanity has transpired.

On the one hand, I think it is unlikely that human minds will ever be able to be "uploaded" to computer systems in order to live virtually in perpetuity. We are not simply the information content of our brains, but also the bodies that live out and process our experiences, interact with the world, remember and forget. On the other hand, the pursuit of uploading as a method to achieve virtual immortality is a pursuit that takes time, resources, and political and social energy away from the pressing problems of the present. It is not that we should not look to solve the problems of the future, but rather that chasing after a highly improbably dream puts our focus in the wrong direction and ignores both the present and future needs of humanity in interaction with our technology.

There is nothing inherently problematic with wanting to live a long and healthy

life. But the quest for technoimmortality raises ethical issues regarding power, resource allocation, and other environmental and social considerations. In the first place, the development of life extension technologies, the healthcare resources to utilize them, and the technologies needed to upload the informational content of the brain are all very expensive to develop and utilize. It is reasonable to assume that advanced life extension and uploading technologies would be prohibitively expensive for the vast majority of the world's population, even if for affluent Westerners they were "almost free."[271] Nor does the movement for radical life extension and technoimmortality take into account the vast environmental resources required to develop and sustain such a program on a broad scale.[272]

Further, while normally there is a passage of power from one generation to the next in business, politics, and family life, the sudden availability of virtual consciousness

---

[271] Kurzweil, *The Singularity is Near*, 469. As I have argued elsewhere, "It is hard to avoid the conclusion that dramatic disparities in power would emerge in the skewed demographics that would follow the development of RLE (radical life extension) technologies. The assumption may reasonably be made that RLE would not arrive as a universal benison coextensive with the human race. That is, like all new and costly technologies - not least, in and around healthcare - the trickledown from wealthy early adopters to the rest would be gradual and would not go far. With good reason, it is understood in the political and economic spheres that people in power will go to great lengths to maintain and increase it." Amy Michelle DeBaets and Nigel M. de S. Cameron, "Be Careful What You Wish For? Radical Life Extension coram Deo: A Reformed Protestant Perspective," in *Religion and the Implications of Radical Life Extension*, ed. Derek F. Maher and Calvin Mercer, (New York: Palgrave Macmillan, 2009), 43.

[272] The availability of life extension technologies and uploading would likely further exacerbate environmental impacts, such as overcrowding, urban sprawl, and natural resource depletion. "Despite the declining fertility levels projected over 2005-2050 the world population is expected to reach 9.1 billion according to the medium variant and will still be adding 34 million persons annually by mid-century." This does not take into account significant gains in lifespan. *World Population Prospects, 2004 Revision, Volume III: The Analytical Report*, United Nations, 2005, p. xiv.

could produce something like a "tyranny of the baby boomers," in which legislators and supreme court justices, business executives, and even professors could retain their positions in perpetuity, stifling both the opportunities for advancement and influence of the generations behind them. Moravec's scenario, however, is far more disturbing. In his vision of the future, humanity is annihilated in the quest for the intelligence of the universe. The inevitable deaths of billions of people are greeted with a virtual shrug. The challenge of political quietism will be explored further below.

3. Evolution and the "Law of Accelerating Returns"

A progressive theory of evolution, whether in biology, culture, or technology, is a widespread, though not universal, problem among futurist thinkers. As noted in chapter 2, Ray Kurzweil derives his concept of the impending Singularity from the idea that cultural and technological evolution (growing out of biological evolution) is moving in a particular trajectory, toward a particular goal, and doing so at an ever-accelerating pace. Without this understanding of the necessary and inevitable trajectory of history, the threat (or promise) of the Singularity coming in the next 30-50 years is substantially reduced. But it is important to analyze the assumptions about the nature and function of evolution that lead Kurzweil and Moravec, among others, to this conclusion.

The idea of a Singularity, or the impending radical change in human life envisioned by Kurzweil and Moravec, depends upon several problematic assumptions. First, it requires a serious misunderstanding of the nature of "evolutionary processes," particularly in Kurzweil's work. His entire idea of the Singularity rests on the assumptions that

evolution has a particular trajectory and momentum with a goal of the development of "intelligence," that evolutionary processes have a built-in momentum toward this goal, and that the evolution of "intelligences" like technology will develop naturally as part of these processes of evolution. He claims that evolution is ordered toward the development of ever-increasing "order."[273] This order then increases exponentially as evolution builds upon itself and the "valuable products of the process" increase at an exponential pace.[274]

Not only is the idea that evolution moves toward greater "order" simply a misreading of evolutionary science, the concept of the "valuable products of the process" provides a moral valuation that is not present in evolution itself. Evolution does not move toward greater intelligence and order, as Kurzweil believes - it does not move "toward" any goal at all. It simply moves in the struggle for survival, and biological evolution (as well as technological evolution) adapts through selection within a particular given, yet changing, ecology. That which survives within the given ecology reproduces itself (with mutations), and as an ecosystem changes, those mutations that survive further reproduce. The idea that technological evolution is inevitable is a substantive misreading of history that fails to take into account both the contingencies of human existence and the very real problems that can and do result with the technological development. In the "more is better" paradigm, he resists nonlinear and non-progressive readings of history and largely ignores what Nick Bostrom describes as the "existential risks" of late-modern

[273] Kurzweil, *The Singularity is Near*, 32-33.
[274] Ibid., 33.

technological development.[275]

Kurzweil misrepresents evolution as a form of development, which leads to his moral overvaluation of technology as the pinnacle of history that will make everything better. This then leads to the political quietism described later in this chapter, but it also leads technologists to embrace bad science and even the manipulation of science and scientific claims for their own ends.

The problematic nature of this understanding of evolution is difficult to overstate, yet it has remained largely uncriticized in responses to Kurzweil's work. Kurzweil does not limit his understanding of "evolution" to biological evolution by natural selection, but rather he incorporates it into a larger framework that includes biological evolution but envisions it as part of a larger trajectory of history. Stephen Jay Gould, one of the foremost evolutionary biologists of the 20th century, resoundingly and repeatedly criticized the idea of a progressive and inevitable evolutionary teleology throughout his work, saying,

> In short, graspers for progress have looked exclusively at the history of the most complex organism through time - a myopic focus on extreme values only - and have used the increasing complexity of the most complex as a false surrogate for progress of the whole…But this argument is illogical and has always disturbed the most critical consumers… I do not challenge the statement that the most complex creature has tended to increase in elaboration over time, but I fervently deny that this limited little fact can provide an argument for general progress as a defining thrust of life's history. Such a grandiose claim represents a ludicrous case of the tail wagging the dog, or the invalid elevation of a small and epiphenomenal consequence into a major and controlling cause.[276]

---

[275] Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Journal of Evolution and Technology* 9, no. 1 (2002).
[276] Stephen Jay Gould, *Full House: The Spread of Excellence from Plato to Darwin*, (New York: Harmony Books, 1996), 168-169.

Gould takes on the idea that intelligence has more than a cursory place in the grand story of the universe, claiming instead that "The outstanding feature of life's history has been the stability of its bacterial mode over billions of years."[277]

Philippe Verdoux has recently analyzed this tendency in futurist thought, including Kurzweil's, and has sought to rid futurism from this discredited Enlightenment narrative. Verdoux argues that "the progressionist conception of history as 'a record of improvement in the conditions of human life' is highly problematic, both empirically and methodologically…(while) most transhumanists today accept progress as a 'central dogma' of their technocentric worldviews."[278] Things don't get better simply because we have the latest and greatest technology. Mistaking evolutionary history for the narrative of the development of intelligence in the quest for an artificial superintelligence is one that historian David Noble describes as a religious phenomenon with an "eschatological vision."[279]

Antje Jackelen critiques this understanding of evolution from a theological perspective, making arguments similar to Gould's: "Biologically, evolution operates by means of mutation, selection, and adaptation. In terms of biology, there are no values attached to these concepts. Very often, however, and fairly uncritically, we tend to attach certain values to evolutionary processes. Thus, we talk in terms of adaptation but often

---

[277] Ibid., 175.
[278] Philippe Verdoux, "Transhumanism, Progress, and the Future," *Journal of Evolution and Technology* 20, no. 2 (2009): 50.
[279] David Noble, *The Religion of Technology: The Divinity of Man and the Spirit of Invention*, (New York: Alfred A. Knopf, 1997), 164.

understand it in terms of improvement."[280] Likewise philosopher Mary Midgley severely critiques the "predictions of the indefinitely increasing future glory of the human race…(that) claim scientific status, not just because they appear in scientific books, but also because they take claim to take their start from the biological Theory of Evolution."[281]

Not only does this misreading of both the nature and function of evolution constitute poor science and lead to an apathetic acceptance of the status quo, it strongly overvalues the contributions of technology and (a narrow understanding of) intelligence. This focuses technological development on only one particular set of functions and values in human life and ignores the richness and diversity of many ways of being human, utilizing technology, and being intelligent. When one assumes a single trajectory of history, it leads to a misreading of past, present, and future by ignoring all the various aspects of life that do not fit into that particular reading. It ignores the mistakes of history, the contingencies, and the possibilities that human choices have. It disallows many forms of creativity and reads only the trajectory of the victors and the wealthy in a technocratic society. And it overlooks or fails to take seriously the problems that technology can create, as I will explore further later in this chapter.

4. Determinism and Political Quietism

---

[280] Antje Jackelen, "The Image of God as Techno Sapiens," *Zygon* 37, no. 2 (June 2002): 296.
[281] Mary Midgley, *Science as Salvation: A Modern Myth and its Meaning*, (New York: Routledge, 1992), 147.

One of the most unfortunate consequences that arises from futurist positions that rely on concepts of inevitable technological trajectories of development (e.g., Kurzweil, Moravec) or simple crass determinism (e.g., Minsky, Hall) is the logical next step of political quietism and inaction. For if there is no possibility for meaningful political, social, or other action with regard to the development of technology, there is no reason to try to direct the future interactions of humanity with our technology - everything is either going to happen as a natural product of "evolution" or is a function of determinism and chance, and neither leaves room for choices that alleviate some of the crushing problems for human life that exist currently or might arise with future technologies. There is thus neither concern for nor action to address present *or* future problems surrounding the development of technologies in general, and humanoid robotics in particular. The human race will be superceded, and there is nothing to be done about it but accept it and watch the beauty of technology unfold, even as humanity is wiped out. There is no way to choose what we will actually develop in our technologies, how they will be utilized, and what limits should be placed on technological development in certain directions - everything that can be developed not only should be but will be developed along the trajectories started from before the dawn of humanity. This kind of deterministic thought lulls us into accepting the status quo in both the present and the future and provides an easy escape from genuine responsibility for our individual and collective choices and actions.

Kurzweil seems to sincerely believe that the LoAR will lead quite naturally to a paradaisical future of superintelligence spreading throughout the universe. His

understanding is that life beyond the Singularity is fundamentally unknowable, although the way that he reads technological development as an unproblematically good thing leads him to project that the future of intelligence on the other side of the Singularity will be a substantial expansion of the wonders of technology that already exist. He envisions the future of humanity as one that leads to technoimmortality, wealth, and pleasure, even as jobs and bodies cease to exist. With advanced technologies, diseases will be eradicated and wealth will flow to all. This naivete about the realities of human nature will be explored more below, but it is his understanding that this will happen simply because evolution will make it so without any intentional intervention. That assumption encourages inaction to transform the injustices that currently characterize human societies.

Moravec's understanding of the future of technological evolution likewise encourages political quietism and acceptance of the status quo, as technology will develop a grand superintelligence that expands throughout the universe and overtakes humanity. It is his hope that he (and some other humans) might be able to be incorporated into the technology that develops and that the coming machine superintelligence will feel some charitable nostalgia for their human forebears and incorporate us into their systems in some capacity even as our time in history comes to a close. Of course, the vast majority of human beings will not be maintained in this fashion as humanity is eclipsed, and Moravec writes off the end of humanity with relative ease and unconcern.

Minsky's determinism arises from a different source than that of Kurzweil or Moravec. His theory of mind (which is useful in its emergentist understanding) falls into the trap of believing "made up of = no more than." That is to say, where he understands

human minds to be made up of a vast network of interconnected, yet on their own, nonthinking and mechanistic, parts, he does not allow for meaningful choice as one of the potential emergent properties that may arise when a mind becomes conscious of itself as a thinking, emotional subject. He reduces all of existence to "determinism and chance" and does not allow for any meaningful choice to exist, even while he allows for the illusion of choice in conscious subjects.[282] This determinism arising from an otherwise interesting and useful theory of mind is not only reductionistic; it is ethically problematic, as it leads to the same kinds of acceptance of the status quo and inevitable trajectories of technology and politics that Moravec and Kurzweil fall into.

However, once the idea of inevitable evolutionary determinism is rejected, then human responsibility for technology re-emerges. There is no evolutionary reason to believe that things are inevitably moving in a single direction, so political and social action become meaningful once again. Understanding evolution not as a trajectory toward intelligence but as adaptation and selection for survival within contexts of niche ecologies within a broader, interdependent ecosystem provides a different perspective from which to understand the choices we make in technology and the long-ranging effects of those choices.

5. Political Naiveté / Technology Solving Social Problems

Yet another key ethical problem that arises within robotic futurist thought that is tied to understandings of the inevitability of technological development and political

[282] Marvin Minsky, *The Society of Mind*, (New York: Simon and Schuster, 1986), 306.

apathy and quietism is the reliance on technology, rather than political, economic, and social choices, to solve major problems within human society, such as poverty, hunger, lack of medical care, clean water, war, etc. This arises in much futurist thought, which tends either toward the utopian or dystopian, but is particularly prevalent among roboticists.[283] Popular imagination in science fiction and futurist literature drives toward the idea that our intelligent machine creations will either save or damn us - most often, they will take over the world and destroy us in the process, rising up against their human masters.[284] Robotic futurists tend to be among the most optimistic in believing that technology will solve major social problems. For instance, Kurzweil claims that

> Emerging technologies will provide the means of providing and storing clean and renewable energy, removing toxins and pathogens from our bodies and the environment, and providing the knowledge and wealth to overcome hunger and poverty…These technologies will create extraordinary wealth, thereby overcoming poverty and enabling us to provide for all of our material needs by transforming inexpensive raw materials and information into any type of product.[285]

These assumptions about the possibilities for technology to solve problems must be unpacked and their problematic ethical implications explored. Kurzweil claims that emerging technologies "will" provide solutions to social problems, but they cannot do this unless they are intentionally developed, broadly accessible, and utilized without causing further social or environmental problems, and he does not take these political realities

---

[283] Robert Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*, (New York: Oxford University Press, 2010).
[284] This imaginary has a long history in science fiction, ranging from Asimov's robots, who seek to save humanity from ourselves by enslaving us to the Cylons of *Battlestar Galactica,* who rise up against the humans who created them and annihilate human civilization across several planets.
[285] Kurzweil, *The Singularity is Near*, 371-372, 396-397.

seriously. First, the funding and will must be present for technologies to be developed that create the conditions of the possibility for major social problems to be solved. Solutions do not create themselves. They have to be chosen, nurtured, developed, and distributed. There must be policy structures in place that encourage the development and mass distribution of useful technologies. There must not be policy structures that encourage detrimental development or hoarding of intellectual property in such a way that innovation is stifled, as in Michael Heller's conception of the anti-commons, when too many owners effectively block one another.[286] Once beneficial technologies are developed, there must be institutions and incentives that ensure that they are widely and fairly distributed to the public as a whole and do not accrue only to the wealthiest members of society. Finally, they must actually be used in such a way to to provide the promised benefits to a broad group of people without causing harm to others.

Kurzweil, among others, likewise assumes that emerging technologies will create "extraordinary wealth" that overcomes the entrenched problem of global poverty. He is right that technologies do tend to increase the overall wealth available, but not that the simple emergence of new technologies will substantially reduce poverty on a broad scale. On the contrary, the vast majority of the gains in wealth in the past generation have gone to a small number of already-wealthy, technologically-savvy people, while the majority of persons (including, but not only, in the United States) have seen their incomes fall and their job prospects diminish. Far from eliminating poverty, technological innovations like

[286] Michael Heller, *The Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation, and Costs Lives*, (New York: Basic Books, 2008), 2.

robotics have increased productivity by replacing jobs traditionally done by factory and service workers, thus leaving those workers without technological educations without employment. This has been played out across a number of industry sectors, particularly in manufacturing. There is no reason to believe that this trend will somehow reverse itself without intervention, as in the mass retraining of employees to new sectors of the economy and educating young workers in areas that are less likely to become automated through the use of robots and other systems.

Finally, Kurzweil's hope for the technological future rests on unsound assumptions about the availability of materials used in technological development and distribution. While nanoscale particles are being developed and utilized in industrial products, the hope that common, cheap materials can be altered at the molecular level to become any needed product quickly and cheaply remains a distant dream. For the foreseeable future, we will have to make do with existing materials, and this means that shortages of rare and costly materials are likely to arise, along with attendant political and social conflicts over the control of those materials, as has been seen with gold, diamonds, rare earth minerals used in computers, and other highly valued products.

Although new technologies can contribute to solving social challenges, there is no reason to believe that they can solve all existing problems. Indeed, they can even create new difficulties. Philippe Verdoux's analysis of the tendency among optimistic futurists to assume a progressive narrative of history shows how that view effectively blinds them to the social, political, economic, and other problems that emerging technologies actually

create.[287] Nick Bostrom's recent futurist work (though not focused on robotics development) highlights the potential "existential risks" posed by various new technological developments. These existential risks arise from the relatively greater power that arises within 21st century technologies - the risks are such that they have the potential to annihilate human life on earth.[288] The real risks of technology - whether existential, social, or political - must be taken as seriously as the benefits in order to develop and utilize technology ethically. In this, there is reason to hope that humanoid robots can be developed and used ethically, but it is likewise a cause for vigilance, even skepticism, that technology will solve social problems.

6. Politics and Privilege Without Recognition

Along with the political naiveté that often accompanies robotic futurist thought - the bright and shiny futures in which everyone is wealthy, happy, and long-lived - lies a serious lack of recognition of the various forms of political, social, economic, educational, gender, racial, ability and national privilege that are currently bound up with the practice of futurism itself. That is to say, futurism, and the technology policy and development that are connected to it, are dominated by older, white, Western, educated, able-bodied, upper middle-class men who fail to reflect upon, much less understand, the implications of their social locations on their decisions. For instance, when Moravec is undistressed by the annihilation of humanity by our robotic descendants and believes that he will enjoy a

[287] Verdoux, "Transhumanism, Progress, and the Future."
[288] Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards."

"comfortable retirement" as a cyborg or uploaded mind, he betrays the assumptions of his position of extreme privilege. He believes that he will be considered sufficiently valuable, as a historical relic, to be maintained somewhat indefinitely by his "mind children." He does not stop, even for a moment, to consider the human devastation and suffering that his hoped-for revolution in robotics could cause. Moravec remains utterly callous toward those whose technological achievements do not match his own and who would be the casualties of technological development. Kurzweil likewise fails to imagine anything but success and wealth accompanying radical technological development. Where he envisions the end of work - human labor replaced by robots - he imagines that everyone will enjoy wealth, leisure, and happiness, and not the possible misery of mass unemployment and hunger that often occurs when human labor is replaced by robots in manufacturing and related industries.

What is needed, instead, is ethical imagination, enabling consideration of the various ways that technological development impacts different people. It is crucial to begin to have a global perspective on the future, which requires both a less optimistic understanding of history and a willingness to engage with people whose values and visions of the future are very different. Opening up public discussion regarding the values built into technology requires attention to the ongoing effects of colonialism and the ways in which current trends in technology development create new forms of colonization.[289] The

[289] One example of this would be attention to Moravec's story of the arrival of Westerners in Papua New Guinea, which he describes in terms that indicate that the local inhabitants were benighted, superstitious, and otherwise uncivilized and in need of the advanced technologies of the West. Moravec, *Robot*, 3.

discussion requires reflecting on the impacts of gender on both historical and current developments in robotics and asking about the funding, design, and development of robots.[290] It requires difficult engagement with the culture of robotics itself as it has been shaped and patterned by certain perspectives of power and privilege. This discussion is not an all-or-nothing affair, solely a matter of one aspect or another, but rather the development of an ongoing pattern of reflection on socio-political location and self-aware inquiry regarding the values built into technologies.

Constructive Anthropological Possibilities in Robotic Futurism

I will now turn to some of the most constructive points of reflection within futurism that offer unique and helpful anthropological and ethical possibilities for thinking about the future of human engagement with humanoid machines. As we begin to think about the necessary aspects of personhood that can be developed in robots, these interrelated concepts can be brought together to begin developing a working model of human and robotic anthropology and can be integrated into an overall research program. These possibilities include a useful theory of mind, including a recognition of the role of

[290] One of the most encouraging aspects of the current generation of robotics students is the rapidly increasing number of women in a field that has historically been one of the most male-dominated. Women in robotics, like Cynthia Breazeal and her counterparts, have begun to ask questions about psychology, sociality, and forms of embodiment that had previously been unconsidered. There remains quite a bit more work to do, as even the current generation of women in robotics is trained to think, design, and reflect in the patterns of their male predecessors, but there are some hopeful signs of a shift in the culture of robotics development.

emotions and ethics in personal life, that gives rise to the embodiment, sociality, and situatedness of humanoid robots.

1. Theory of Mind - Emergence, Emotion, and Ethics

While I find that Minsky's simplistic hard determinist understanding of the world to be entirely unhelpful, he offers an emergentist theory of mind that can be useful in thinking about the basis of embodied robotic sentience. In his theory, that which we think of as "mind" arises from an embodied, interconnected web of non-thinking parts. Each individual processing unit is, in itself, simply a small mechanism that functions without thought, but when groups of these mechanisms interact with each other, conscious thought of various varieties can emerge. This need not be taken to the extreme position that Minsky takes that since the individual pieces are non-thinking and appear to be determined by necessity and chance, therefore the emergent consciousness embodied in networks of those pieces can be reduced to the same properties. A robust theory of emergence ought to claim otherwise - that the whole is actually greater than the sum of its parts, and that the interconnectedness of the pieces in specific kinds of configurations can give rise to properties not present in the individual parts. This is true of the bases of life itself - on their own, the individual bits of matter that make up living beings are not, in themselves, alive. Yet when they are joined in certain patterns of configuration, something greater - life - emerges from the non-living pieces, and living things clearly have very different properties (and capacities) than the individual pieces. Philip Clayton has argued for a strong emergentist theory of mind - like Minsky, he finds that embodied minds have

properties that are unique to the complex configurations of simpler processes, but he instead allows for "top-down causation": that the emergent properties of the whole affect the ways that the parts interact with each other such that those interactions could not be predicted from the properties of the parts.[291]

Minsky's theory of mind also helpfully intersects various types of processes and their interrelations with one another. He describes human minds as "emotion machines," indicating that our rationality, actions, and modes of thought are influenced by a complex set of smaller processes that cannot be abstracted or divorced from one another. Our minds are embodied in interactions of simpler physical processes, and these processes are fundamentally inseparable. Our rational functions are inextricably tied to our emotional functions, self-preservation, sense of time, etc.

If we take the best aspects of Minsky's theory of mind and include a stronger sense of emergent properties from Clayton's theory, then there exists a solid case for a working theory of mind that could apply to both human and robotic persons. First, minds are necessarily embodied - the processes of interactions are bound up with the networks and interactions between them. Second, conscious minds are emergent properties of smaller and simpler systems, and the workings of the whole system are not immediately derivable from the functions of the subsystems. Third, this then allows for the genuine possibility of emergent minds arising from sufficiently complex (and specific) networks of interaction between systems, regardless of the substrate in which they are embodied. It is

[291] Philip Clayton, *Mind and Emergence*, (New York: Oxford University Press, 2004), 52.

not necessarily the case that such a mind will necessarily emerge - the networks of interaction are sufficiently complex that we do not yet understand what it is that gives rise to human consciousness - but rather that the emergence of mind from a machine substrate is not inherently impossible. It may be the case that such consciousness cannot arise; there could be something as yet unknown in carbon-based life systems that gives rise to emergent consciousnesses in ways that other substrates cannot, but this emergentist theory of mind allows for the question to remain open.

Minsky's theory of mind, along with that of other developmental theorists, moves far away from the dualist-rationalist conceptions of human intelligence that are common among robotic futurists. Arguing for the necessity of different types of systems - including, and perhaps especially, emotion - within the full spectrum of human thought requires a break from the analytical and symbolic-manipulation oriented thought that has characterized much historic work on the development of "artificial intelligence." It is not simply a matter of developing the "computational capacity of the human brain" that could make a robot a "who" rather than solely a "what."[292] Machines can now perform many computational processes that far exceed those of human capabilities, yet there is so much more to human life and functioning than simple computational capacity.

One can use a framework like Howard Gardner's theory of multiple intelligences in order to understand this shift in understanding toward a multifaceted view of what "intelligence" requires, either in humans or robots. Beginning in 1983, developmental psychologist Gardner led a movement away from traditional understandings of intelligence

[292] Kurzweil, *The Singularity is Near*, Chapter 3.

(particularly IQ) as being only a matter of one thing - logical-analytical skills - and instead formulated a theory that there exist in human beings roughly eight or nine kinds of intelligence.[293] These include linguistic, musical, kinesthetic, and interpersonal intelligences alongside traditional logical-mathematical intelligence. While his work is not uncriticized and has evolved over time, it is useful in thinking about the role of different types of processes and systems within human embodied thought to understand the full range of "intelligence" in human life that would need to be embodied in a fully humanoid robot. This theory can be useful to understand intelligence as a whole range of skills, talents, and interests through which we relate to the world around us, and not simply the ability to solve certain types of problems.[294]

This aspect of a working theory of mind calls needed attention to the emotional and creative aspects of human life and intelligence. These, more than the ability to manipulate symbols, inform and help us navigate the world around us and the problems that embodied human life presents. We do not interact with the world primarily through logical problem-solving.[295] Our first and possibly most important interactions with the world come through mediated bodily interactions and interpersonal relationships. It is

---

[293] Howard Gardner, *Frames of Mind: The Theory of Multiple Intelligences*, 3rd edition, (New York: Basic Books, 2011), 64-66. His original version of the theory included seven intelligences; later revisions have eight, and he is reportedly considering a ninth.

[294] This is true regardless of what one makes of Gardner's theory in relation to its primary application in education.

[295] In making this claim, I have no wish to denigrate the importance of logical problem-solving, as it has many very important applications in human life and is a critical aspect of general intelligence. I mean only to claim that this form of intelligence does not stand alone and may not even be the primary form of intelligence needed to navigate the world "humanly." But it has been overvalued by roboticists and others, at least partially because of how it correlates to their own primary intelligences and the problems they are seeking to solve.

through the bonds that we develop with one another early in life that we are later able to consider the problems of logic. We are physical creatures with bodily needs and emotional interactions, and these must not be left behind in the quest for a highly intelligent robot. As another physical being in the world, a humanoid robot must also learn to navigate its needs for food (electricity), shelter, and relationships.

Understanding the role of emotions in human life also highlights the need to study the ways that humans interact with machines, both positively and negatively. It is encouraging that researchers at places like MIT are studying what factors influence what types of machines we prefer to interact with and why.[296] Learning why people respond emotionally to different aspects of a robot's physicality, voice, and modes of interaction can be useful in developing useful and non-threatening robots. For instance, making a robot sweet and helpful, not too bright, and not too human, tends to encourage human interaction with the robot.[297] This can be used for a variety of purposes, including manipulation of humans interacting with the robot, but it can also be used to design robots that people want to engage with and alleviate some of the fears that people have about robots.

Also important in the development of humanoid robots is how we structure their

---

[296] This is a key insight of Cynthia Breazeal and her students, who have worked on sociality, cuteness, facial expressivity, and facial structures that influence whether people choose to interact with a robot. One recent creation is "Boxie," which is simply a cardboard box on wheels with wide-set eyes and a smiling mouth. Boxie has an internal video camera and relates to people much like a lost toddler, speaking in a high-pitched voice, asking for help, and recording people's responses. The vast majority of people respond positively to its helplessness. http://boxie.media.mit.edu/

[297] The design of the robot WALL-E in the Pixar film of the same name is archetypical for generating positive interactions. WALL-E looks kind, but not too human.

modes of behavior and interaction with humans - to make them "ethical" subjects as well as objects. I believe that it is critical to consider what decision-making powers are given over to machines by humans, and it seems that there are constructive possibilities for designing programs within robots that influence their behavior for the better. But all such choices must be based on particular ethical systems, which remain hotly contested in the realm of human choices. For instance, one would build in very different parameters and decision trees if one wanted the machine to be a utilitarian, rather than a deontologist. The factors that would be considered in the choice of action are different, and the resulting actions would likewise vary.

It has been proposed that robots be designed to obey something like Asimov's Three Laws of Robotics, though functionally, this is far more difficult than it first appears. Following the First Law (a robot may not harm a human being or allow a human to come to harm) alone is exceedingly complex. There are two problems of knowledge - what would cause harm to a human being and the multiple impacts of any given action on the wide variety of human beings who could be affected by the action. These two problems of knowledge then lead to a problem of action - nearly any action that one takes has at least the possibility of harming someone. Finally, the First Law directly conflicts with the duties of warfighting robots, who are not directed to avoid harm but to (carefully and proportionally) inflict it. Many of the same problems arise if one chooses to build a robot with a utilitarian ethical calculus - the knowledge of what helps and what harms any number of human beings (or animals, the environment, etc.) is always fallible and probabilistic, so determining action with more than the most vague approximation is

difficult. It is possible to theoretically build rule-based models of robotic action that do not have this level of ethical calculus, but rather prescribe boundaries of behavior within which the robot can act. In this case, the robot could be programmed to not strike a living being and be built with the appropriate actuators to enable this. Within this single limit remains a wide range of action options, and the robot is not left searching for the best ethical action - it does its job and is programmed with a basic system to avoid the most likely possible harms.[298] Much more work needs to be done, both in theory and practice in the design of such robots, but the fact that building ethics into the design and action of robots is becoming increasingly important in the field is a positive step.

The emergentist theory of mind, including its emotional and ethical component, is necessarily materially embodied. This mind, in order to develop, must exist in concrete relationships and have an understanding of its situatedness in culture, history, and place. It is using this theory of mind that the most socially complex humanoid robots have been developed, and it is to this embodied situatedness that I now turn.

2. Embodiment

Much has been said up to this point regarding the place of embodiment in a working theory of mind, of emotional relations between human and machine, and the forms of intelligence at work in fully humanoid robotics development. Insistence from

[298] Again, this would not apply in the case of military robots, which require an entirely different set of "ethical rules." This follows Brooks' design of robots that respond directly to changes in their environments, rather than Hall's, which proposes the cut through the problem using evolutionary biology.

feminists and others that a robust understanding of the importance of concrete embodiment in the nature and functioning of human life is well-grounded, and some current-generation roboticists are delving into why the facts of embodiment matter so much. While an older generation emphasized "intelligence" as the key component to human life and cognition as the essence of personhood, a new generation, starting with Rodney Brooks and flourishing with Cynthia Breazeal, is investigating the impacts that differing forms of embodiment in robots have on human-machine interactions in the real world.

"Embodiment" as a concept is critical to understanding what it means to be a human being in the world. Lakoff and Johnson have shown the myriad ways in which the particularities of human embodiment in the world directly impact the form and functioning of our rationality.[299] "Embodiment matters," one may say, but it ought not be either romanticized or overdramatized. On the one hand, many aspects of human life encourage an optimistic understanding of the importance of embodiment. Life in the flesh can be a truly pleasurable experience. One cannot begin to understand the heights of human joy - the delights of wonderful food, beautiful art, passionate sex, etc. - apart from the existence of human beings as bodies. It is an ethical, as well as an aesthetic, good to encourage and support the flourishing of human beings as bodies. On the other hand, one's human existence as a body is also one of inevitable suffering, and we ought not overlook this in a rush to explain the amazing importance of the body as our mode of being in the world. Simply put, it is not only because embodiment is wonderful that it is important; it is also

[299] Lakoff and Johnson, *Philosophy in the Flesh*, chapters 3 & 6.

because life as a body is painful and messy that it matters so much. I do not cherish my sufferings, nor do I believe that there must be some hidden meaning in them, but they are a necessary, if messy, part of who I am. We exist as bodies, with both the joys and sufferings that go along with that, and there is not an escape into something other than embodiment that can provide us with the former but not the latter.

Likewise, "artificial intelligences" that do not possess a comparable experience in the concrete, embodied world may be able to process vast amounts of information fed into them, but they are unable to achieve anything like a humanoid, or even concretely intelligent, existence. Just as it is our embodiment, and not only our minds, that makes human life what it is and gives it value. As such, humanoid robots must also be embodied minds in order to have the possibility of existence as "persons."

3. Sociality

Another aspect of humanoid robots that indicates movement in the direction of the fully humanoid is the development of "sociable" robots. Just as human life is a messy mix of the good, bad, and other aspects of embodiment and emotion, so also human life exists in a web of relationships. It is through our engagements with others that we become ourselves - become human. As with embodiment, sociality is not just a rosy view of the ways in which we interact with one another - it has its own forms of messiness and suffering. Human sociality is formed from the very beginnings of life. One of the most unique things about humans as a species is the extraordinary length of time that we spend maturing and the depths of dependence that are inherent in the human maturation process.

We emerge from the womb as totally dependent creatures - we cannot walk, talk, find our own food, or defend ourselves for a very long time after birth, so we need other humans to care for, feed, and protect us while we mature.[300] Thus we are marked by sociality from the very beginning. There is, quite literally, no such thing as a self-made man - our dependencies early in life provide not only for our survival during the early years, but also give us the tools to survive and succeed as we mature. Even the hermit in the woods had someone to feed, shelter, and nurture him to adulthood.

Our sociality continues to shape us throughout our lives. Humans generally live in societies of some kind in order to make the best use of our brains for our own protection and nurturance. We depend on one another for delivering what we need for the complexities of life as creatures with few natural defenses but big brains. If we are fortunate to reach old age, we enter a new dependency, and if we encounter setbacks, illnesses, disabilities, or the like along the way, we often have dependencies in addition to those that all humans face. All of this is to say that our sociality - our relationships with other human (and non-human) beings are integral to who we are and what makes us persons. Our striving to develop "intelligence" in our robots must include the intelligence to navigate relationships, dependencies, collaborations, and competition in a social world.

Sociality is often not a glamorous thing. Contrary to the understandings of many communitarians,[301] life in human relationships is not only nurturing and capacitating but

---

[300] Alasdair MacIntyre, *Dependent Rational Animals: Why Human Beings Need the Virtues*, (Peru, IL: Open Court, 1999).

[301] MacIntyre is probably the most guilty of this in his writings, claiming community as inherently developing virtue.

also damaging, hostile, and filled with greed and rejection. Because we cannot effectively live alone, and because we live out our lives together in ways that are both nurturing and damaging, we need to think critically about how we design humanoid robots that develop strong, healthy relationships, that contribute to human flourishing and do a minimum of damage along the way. To this end, sociality is critically important, and all the more important to design well.

4. Situatedness

"Situatedness" adds yet another layer to the embodied, emotional, sociable understanding of a full human (or humanoid) existence. It refers to the various ways in which we are not only beings who are bodies, who relate to others in the world, but also have a rootedness in a particular time, place, culture, and environment. It is our being in a real, concrete history. It is in our situatedness that we learn and grow and through which we understand the interdependent relationships that shape us. Situatedness includes physical factors, like a home, neighborhood, city, and nation. It involves trust or neglect, the people and media to which we are exposed, the schools we attend and the students and teachers who are with us, the faith and spirituality we are taught, and the social, political, and class environment in which we live and learn who we are in the world.

Reflection on situatedness allows us to understand the broader systems of human (and non-human) life in which we live. We are not merely individuals (even individuals in relationships), and we shape and are shaped by the particularities of the world around us. We do not exist in a vacuum, but rather must learn to see our actions and reactions as they

stand within larger systems and institutions. It is in the details of life that we discover who we are and who we choose to become. We are, in many senses, constrained by our situatedness; there is no view from the outside of time and place, no Archimedean perspective from which to neutrally and rationally understand the world. We are finite and limited in our knowledge and our reach, and this is reflected in our situatedness. Yet we are not only the situatedness in which we find ourselves. Ethically, we must learn to make meaningful choices within the world we are given, understanding and respecting our constraints while also living into the ethical vision of a better world that could be.

For humanoid robotics development, a concrete understanding of the importance of situatedness relates to the factors that allow different robots to have different experiences, to be differentiated by what they learn, to have the possibility of becoming "persons" by being distinguished from one another in their situatedness, their relationships, their emotions, and their embodiment. We must understand how even robots (that are sufficiently complex and robust) are shaped by what they learn, not unlike pets or humans. This aspect of robotics development stands, in some ways, down the road a bit, but if we consider factors of situatedness in advance, we can build in reflective engagement with a broader environment into the robots that are built and lead them one step closer to being fully humanoid.

Futurism and Theology: Toward a Robotic Anthropology

In this chapter I have sought to highlight some of the key ways in which contemporary robotic futurism can be used in developing an understanding of what it

could mean for a robot to be a person, including several key factors to either include or avoid in developing that understanding. When some of the important anthropological and attending political problems are avoided, new spaces open in which a robust view of a humanoid robot can begin. The picture that emerges is one in which both robotic and human persons are physically embodied with emergent minds, have a strong sense of both emotional and ethical awareness, and are both sociable and concretely situated in their environments. To this picture I now wish to add an additional layer, that provided by current reflections on personhood from Christian theology. This theological analysis will begin by comparing the anthropologies of two contemporary theologians whose work reflects on personhood within the context of technoscientific developments. While neither of them specifically addresses an anthropology of robots, their accounts of personhood can be compared and reframed in order to build on the anthropology developed here. This theological account will provide further context in which to understand the emotional and ethical lives of persons, whether human or robotic, and will deepen the reflections on persons' embodied, social, and situated subjectivity.

Chapter 4: Theological Anthropology for Contemporary Technoscience

This chapter will explore some possible contributions of Christian theological anthropology to a better understanding of what it might mean for a human or a robot to be a "person," in a moral and theological sense. The underlying question will explore a theological distinction of what makes a being a "who" rather than a "what." I will begin by analyzing two competing theological anthropologies that are utilized within current Christian reflection through the work of contemporary theologians who focus on the place of emerging technologies in human life: Philip Hefner and Brent Waters. Within this analysis I will attempt to show what important contributions each thinker provides to a comprehensive picture of theological anthropology, the weaknesses of each position, and how they can be used to balance each other for a stronger account overall by deepening the understanding of emergent personhood as embodied, sociable, and situated.

Philip Hefner and Brent Waters: Two Theological Lenses on the (Human) Person

Contemporary Christian reflection on human personhood in relation to emerging technologies tends to fall into one of two general lenses of understanding. Among those who tend to be more strongly in favor of the development and utilization of technology to enhance human life, the most common lens is that of the "created co-creator," a term

coined and developed by Lutheran theologian of science Philip Hefner.[302] His work is used directly by several of the theologians who have reflected on humanoid robotics, and I will use his work as the primary source material from which to understand and critique that set of positions. Among those who tend to be less enthusiastic about the constructive possibilities of technology, a different lens is used to view the human person, one that I designate the "embodied-finite-mortal creature." Theologians who take versions of this view have tended to not write on robotics, focusing instead on technologies, like genetics, that change existing human persons instead of imagining the creation of artificial persons. While this constellation of positions has no founding figure, it may be viewed through the lens of one of its strongest representatives, Congregationalist ethicist Brent Waters.[303] Each position illuminates some significant features of Christian theological reflection on created personhood and obscures or neglects others. When their positions are upheld in concert with one another, rather than in contradiction with each other, these two lenses form a more genuinely three-dimensional theological picture of the person. This picture, while greatly improved by reading Hefner and Waters together, can be further strengthened by framing their arguments and setting them in context.

Hefner's work in developing the theory of the human as "created co-creator" is

---

[302] Philip Hefner, *The Human Factor: Evolution, Culture, and Religion*, (Minneapolis: Augsburg Fortress, 1993), and *Technology and Human Becoming*, (Minneapolis: Augsburg Fortress, 2003).

[303] Brent Waters, *This Mortal Flesh: Incarnation and Bioethics*, (Grand Rapids: Brazos Press, 2009), *From Human to Posthuman: Christian Theology and Technology in a Postmodern World*, (Burlington, VT: Ashgate Publishing, 2006), and *Reproductive Technology: Towards a Theology of Procreative Stewardship*, (Cleveland, OH: Pilgrim Press, 2001).

rooted in a religious and anthropological naturalism that takes seriously our biocultural evolutionary history as well as human relations to ultimate reality that are sought in religious myth and practice. For Hefner, the essence of the human person is that of freedom for creativity that is an emergent property of our adaptive plasticity that has been bequeathed to us through a deterministic evolutionary system. Waters' work, on the other hand, stems from a theological understanding that is also rooted in our created evolutionary history. In his conception, the primary essence of human existence is not freedom, but rather finitude and mortality. We are embodied creatures who have been brought into existence and called good by God, and it is in our acceptance of our limits that we find our true expressions of freedom to be the persons whom God intends us to be. I will argue that it is most helpful to read these two thinkers together, rather than in opposition to each other. This produces a picture of the human person that is spiritual in our embodiment, both finite and transcendent, good and sinful, free and constrained. To this picture I want to add a caution about attending to difference within claims about "the" human person, an emphasis on our embodiedness that does not romanticize human finitude and vulnerability, and a resistance to overreliance on natural law in identifying the good in human life.

Philip Hefner and the Created Co-Creator

Hefner's theory of theological anthropology begins, as mentioned, with a complete philosophical and religious naturalism. That is, for Hefner, while we may speak of "that which is ultimate," which we call God, humans are creatures that are entirely physical and

natural in being.[304] There is no recourse to a supernatural element of human spiritual life; rather, human spirituality and relation to God are developed in and through the evolutionary processes of creation and take place in our embodied, physical lives. Hefner seeks to seriously explore what the nature of human persons is in the context of the "biocultural evolutionary model within the physical ecosystem."[305] His is a theological anthropology that is derived from knowledge gained in the sciences, particularly evolutionary biology and cosmology. Because we understand the way the world is through scientific exploration, science then becomes a crucial aspect of doing theology well, and particularly understanding human persons theologically.

Hefner identifies some of the key sources of his thought in developing this theological anthropology. His method of scientific theological exploration and testing of hypotheses comes from Imre Lakatos. From Immanuel Kant he gains his explanation of human "behavioral motivation."[306] His theological sources combine existentialist and process thought, and here he draws primarily on Paul Tillich for his God concept and Alfred North Whitehead for the "conceptual mode" of his theological reflection.[307]

The theological anthropology he develops is intended to serve as a series of hypotheses that can, at least in theory, be tested scientifically in order to be both fruitful for research and potentially falsifiable if that research indicates problems with the hypotheses. He does not offer traditional doctrinal affirmations, but rather "proposes the

[304] Hefner, *The Human Factor*, 32.
[305] Ibid., 28.
[306] Ibid., 34.
[307] Ibid.

theory of the created co-creator," which "represents itself as a hypothesis to be tested."[308] This should be true, in his thought, for all manner of theological affirmations, which should be "accepted, modified, or winnowed out entirely according to whether they do in fact help to account for the performance of faith in thought, worship, community life, and behavior."[309] This method of testing hypotheses to form iteratively better understandings of human life and practice is intended by Hefner to produce a universal narrative account of human nature. Because of our shared history, biology, and place in the planetary ecosystem, Hefner claims that "it seems foolhardy…to entertain the notion that serious thinkers might reject the effort to fashion comprehensive conceptualities and interpretations that can enhance our understanding and behavior."[310] Hefner seeks to hold on strongly to a single scientifically-identifiable metanarrative that encapsulates what is important about human life. Indeed, he believes that his theory of the created co-creator "can serve as a summarizing concept for all that we know about the human being."[311] The theory, he believes, can then be applied in a variety of contexts, understanding human behavior and the trajectories of history through this lens.

> The core of Hefner's proposal is this:
>
> > Human beings are God's created co-creators, whose purpose is to be the agency, acting in freedom, to birth the future that is most wholesome for the nature that has birthed us - the nature that is not only our genetic heritage, but also the entire human community and the evolutionary and ecological reality in which and to which we belong. Exercising this agency is said to be God's will for humans.[312]

---

[308] Ibid. 18.
[309] Ibid., 19.
[310] Ibid., 5.
[311] Ibid., 32.
[312] Ibid., 27.

It is Hefner's intention in developing this theory that humans understand the purposes for which we exist and exercise our moral agency in the world accordingly. He is deeply concerned with human ecological responsibility, and so the connections he draws to our evolutionary history are highlighted in order to connect our identity as humans to that of the rest of creation. Despite Hefner's great appreciation for emerging technologies and his influence on Christian theologians' approval of transformative technologies, his anthropology is intended to put something of a check on this tendency for technological overreaching. He intends his anthropology to show how humans live in connection to other humans and to the nonhuman creation, so technology cannot ever overtake the meaning of human personhood, but must be subjected to it. "Any concept of *Homo faber*, aggressive technological operator, that overlooks these basic qualities [of connectedness and interdependence] is clearly inadequate and even perverse."[313]

> Hefner summarizes the theory in three related claims:
>
> 1. The human being is created by God to be co-creator in the creation that God has brought into being and for which God has purposes.
> 2. The conditioning matrix that has produced the human being - the evolutionary process - is God's process of bringing into being a creature who represents the creation's zone of a new stage of freedom and who therefore is crucial for the emergence of a free creation.
> 3. The freedom that marks the created co-creator and its culture is an instrumentality of God for enabling the creation (consisting of the evolutionary past of genetic and cultural inheritance as well as the contemporary ecosystem) to participate in the intentional fulfillment of God's purposes.[314]

Before analyzing the full context of Hefner's theological proposal, a few comments about the succinct version of the theory can be made. First, the initial emphasis of the theory is

[313] Ibid., 37.
[314] Ibid., 32.

on the createdness of human beings by God in the context of a purposeful and purpose-oriented creation. Humans' co-creating activity is based in and directed toward the creative purposes of God. God is, first and foremost, a creator, who gives the creation a process through which to become what God intends for it, enables the creation in its fulfillment through the creation of the co-creator, and orients the entire creation toward freedom. Humanity's co-creator status is intended to move us toward the continuing development and fulfillment of freedom according to the purposes of God.

The created nature of the human condition is one that is given by God and oriented to God's purposes. For Hefner, "this speaks of the primacy of God and the divine creating activity."[315] It is by God's initiative that humans are created and are given the status as co-creators, for "the term *created* indicates that the human species did not design its own nature or its role in the world."[316] We humans are not our own creators, nor do we design our own purposes. We are given a nature of freedom by God in order to bring about freedom for one another and the rest of God's creation. This derivative, yet specifically directed form of freedom toward creativity is the essence of human nature and human purpose. Hefner claims human uniqueness based in this divinely-given creative capacity. "Although, clearly, the co-creator has no equality with God the creator, inasmuch as the former is a contingent creature and dependent upon the creator, nevertheless the very use of the word *creator*, even in a derivative sense, establishes a distinctive quality for humans. Furthermore, the word points to the essential character of

[315] Ibid., 36.
[316] Ibid, 36, emphasis original.

the human: to be a kind of creator."[317] All else that may be said of human life, according to Hefner, can be read in this light. Anything that might be said of traditional theological claims, such as the incarnation in Jesus, the sinfulness of humanity, the need for redemption, or the covenantal relationship between God and humanity, must be read first through the lens of this divine creativity and its origination of human co-creativity.

In our createdness and origin in the natural evolutionary process, we are conditioned in a way that both gives us purpose and provides a context in which we may enact the freedom we have been given. Our powers of creativity are derivative and limited, yet they are the heart of who we are and who we are intended to be. We are natural creatures of evolution, yet we are not entirely subject to the whims of evolutionary selection and determinism. Hefner argues instead that it is out of the deterministic system of evolution that human freedom has actually emerged, and that this is necessarily so.[318] The deterministic process of evolution "extends back to the origins of the universe," yet "within this deterministic evolutionary process freedom has emerged."[319] It is not merely that freedom has arisen out of the evolutionary system, but that it is in the particular combination of natural processes and systems of evolution that an entirely new property develops that could not be predicted in advance given the characteristics of the parts of the system.[320] Noting theologian Ted Peters' work on the emergence of human life and freedom, Hefner claims, "that freedom should be produced out of determinism seems

---

[317] Ibid., 39.
[318] Ibid., 30.
[319] Ibid., 30. He speaks here of emergence technically, rather than colloquially.
[320] Philip Clayton, *Mind and Emergence*, (New York: Oxford University Press, 2004).

startling, until we understand that freedom apparently is in the best interest of the deterministic evolutionary system."[321] How freedom can arise from a deterministic system depends on the form and definition of freedom being described. Hefner's understanding of freedom is multifaceted and moves in several directions at once, but it is not simply arbitrary freedom of choice nor an Augustinian necessity flowing through the human will, but is instead oriented toward creativity and responsibility for that creativity. This freedom emerges from within the evolutionary processes because it "is rooted in the genetically controlled adaptive plasticity of the human phenotype."[322] In other words, it is the way that the human brain has developed and adapts to its situation by adapting the surrounding environment that human freedom comes to be. Following John Hick, Hefner claims that this freedom that arises out of the determinism of evolution is intended by God but not given inherently to the creation; it must be freely chosen by the (proto-humans) in the creation. "If indeed it is God's will that there be a free creation, it is necessary not simply that the world be created in a condition of freedom…but that the creation be created so that it could become free of its own choosing."[323]

This emergent freedom that forms the essence of human nature and activity is the basis of all human action within the world, but it is also the basis of ethics and values. The essence of the freedom exhibited in human nature "is that human beings can take

---

[321] Hefner, *The Human Factor*, 30. His anthropological argument and understanding here has almost exactly the same basis as Minsky's, but he takes it in a precisely opposite direction, showing, at least, that evolutionary arguments are flexible and can be used in a variety of ways toward a variety of ends.
[322] Ibid., 30.
[323] Ibid., 44.

deliberative and exploratory action, while at the same time they and they alone must finally take responsibility for that action."[324] This freedom of action within human nature is thus absolutely unavoidable in the context of everyday human life. Our co-creatorship means that we have developed / been given the freedom to create and to take responsibility for what we create and for what we do with what we create. It means that "humans unavoidably face the necessity of both making choices and of constructing stories that contextualize and hence justify those choices."[325] Perhaps ironically, once the choice for freedom has been made within the creation, it becomes a necessary and innate part of that creation and the lives of its creatures. It becomes both essential and unavoidable.[326] It is the unique place of human beings within the cosmos to be the creatures who freely chose our own freedom, thereby making it a necessity within the creation.

The necessity of freedom and the responsibility Hefner claims it entails thus also bring a particular normativity to human free action and creativity. He claims that one can (at least provisionally) derive "ought" from "is" by attending to the natural structures and processes of a thing and seeing the "ought" arise from what makes the thing healthy and whole. This derivation "constitutes a form of natural law theology" within Hefner's thought; we know what should be from the natural functions of a person, species, or ecosystem. Because of the natural derivation of this normativity, Hefner believes that value and ethics are inherent within the created evolutionary order and "need not be

[324] Ibid., 30-31.
[325] Ibid., 38.
[326] Ibid., 97.

imported from the outside."[327] When we claim that something is good, we claim that it is necessary for the survival and flourishing of a creature or system. This derived natural law has significant limitations, though. He does not explain how one could know what was good in a case where the needs of two being come into conflict, or where different goods lead to competing benefits; nor also does he give any indication that he is aware of the problematic power and resource dynamics involved in making this claim about "natural" oughtness.

The oughtness that arises from Hefner's understanding of natural law provides not only an ethical context for decision-making and creative action, it also "may be said to constitute, at least in a tentative way, goals and purposes for human life."[328] So human beings are created by God for the purpose of freedom to be co-creative; this freedom entails a responsibility to be creative in such a way that humans' and creation's freedoms are enhanced in order to be more creative. But simply being creative in order to be free and free in order to be creative does not provide an appropriate ethical grounding; for this, "a concept of wholesomeness is both unavoidable and useful as a criterion governing the behavior of human beings."[329] We may say that this "wholesomeness," or flourishing, is thus the telos by which we judge the outcomes of the creative choices of human beings in freedom. It is from this wholesomeness that Hefner derives his "teleonomic axiom": that from the structures and functional needs of a thing we can derive an understanding of both

[327] Ibid., 31.
[328] Ibid., 40.
[329] Ibid., 42.

the purpose and meaning of that thing.[330]

In arguing that evolution drives toward the creation of free human creatures in order to fulfill the purposes of God in the freedom of all creation, Hefner moves toward claiming a specific teleology inherent in evolution. From the teleonomic axiom arises an understanding of the purposes of the creator in creating both particular beings and in general; we understand not only what is good for us, but our purpose in life by understanding what we are like and what we need to survive and flourish. And this ultimate need and purpose, for humans, is freedom. But a teleology for evolution itself, and the whole of creation, cannot be understood simply by looking at the structures and processes of evolution. Evolution on its own often seems purposeless and pointless; he knows that "it is not possible to assert a teleology for nature, except as an item of faith."[331] While we can read out of natural processes and structures the purposes of particular things, the system as a whole can only be understood to have purposes through a larger picture of faith, and it is this that Hefner believes that Christianity offers: "a message about the nature and destiny of the world" according to God's creative purposes of freedom.[332] The human species, then, is the agent through which the purposes of God are enacted and embodied in the world, working through both human biology and human culture. In this, "the freedom of the human constitutes a zone wherein the entire creation can be said to participate for better or worse in the outcomes of a more highly evolved freedom. The human being is thus the agent for a new level in the creation's freedom, and

[330] Ibid., 40.
[331] Ibid., 39.
[332] Ibid., 13.

an agent that God has raised up."[333]

Within this faith-based concept of evolutionary teleology, then, a concept of progressivism creeps back in to the evolutionary picture. Freedom emerges as the telos of the creation, but only when creatures arrive who can sufficiently choose their own freedom and bring freedom to the rest of creation. Freedom cannot emerge for the bacteria or plants or fish, but only through "the animals that have more complex nervous systems."[334] While humanity may not be the final stage of the evolutionary trajectory toward freedom, the self-reflection on freedom and purpose that co-creatively gives birth to the freedom of the rest of creation emerges only "at the level of *Homo sapiens*."[335] Evolutionary processes have both internal and external goals and ends toward which they move, which can only be believed in faith, but Hefner seems to also believe that the fact of purpose can be gleaned from the existence of evolution itself, though the content of that purpose is grasped only theologically. His understanding has "the implication that the creation, as an evolutionarily constructed system, is a system in movement, and therefore a system that has purposes. In a more technical theological analysis, this means that the creation is fundamentally both eschatological and value-laden or moral."[336] In claiming that movement is necessarily purposive, Hefner elides how this can be known, except as another article of faith, for there is nothing within the system itself that indicates that movement alone entails purpose.

---

[333] Ibid., 48.
[334] Ibid., 31.
[335] Ibid., 45.
[336] Ibid., 46.

Hefner's anthropology that is focused almost entirely on creativity, freedom, and evolution toward ever-greater freedom may seem to have little in common with classical theological affirmations about human persons. The goodness, power, and mercy of God are transformed into a single good of creativity; the human person is intended for freedom and co-creativity, and finitude and fallenness are comparatively de-emphasized. Humans are embodied creatures with purposes intended by God, but it is unclear what place the life or community of faith have, and ethics seems to be reduced to the proper functioning of a system. Hefner does have a place in his work, though, for finitude, frailty, and sinfulness, and they arise from the same place in our biocultural evolution that our freedom does, for "neither the genotypic nor the cultural systems are perfect in their ability to guide and sustain human behavior."[337] As created co-creators, we are free but incomplete and imperfect. We lack the perfections and infinity of the creator and are left with the inadequacies of our evolutionary heritage. This incompleteness and imperfection are the grounds for the actions we understand as sin. There is an "interrelatedness of freedom with a sense of inadequacy and guilt," and this provides the basis for Christian theology's "emphasis on fallibility and finitude."[338] Our finite and fallible knowledge, understanding, and freedom bequeathed to us by the evolutionary processes through which we were created leads to an "unending program of trial and error," just as in evolution itself, that "includes within that unending program the elements of human

[337] Ibid., 135.
[338] Ibid., 136.

awareness, decision, and the accompanying self-aware feedback mechanisms."[339] There is no original sin that is transmitted down through the generations; we are finite, vulnerable creatures who try and often fail in our efforts to embody the creative freedom we have been given, and while we often feel shame at our behavior, the vulnerability and finitude we experience are based in the very freedom we have achieved.[340] Not everything is rosy, though, and there is within human freedom a tremendous capacity for real evil. Our freedom entails responsibility, and we can abuse that responsibility for our own greed, selfishness, and destructiveness. This is, in many ways, most poignant in our use of technology, which Hefner understands as a source of both goodness and evil. There is, he says, "unmistakable grace mediated through the co-creator's technological accomplishments. The demonic is just as vividly present."[341]

Another key theological component that seems to have dropped out of Hefner's theological discourse on the created co-creator is any mention of Jesus or the role of Christ in human life. But Jesus is brought back into Hefner's discourse when he considers a model for the proper shape of human life as created co-creators. Jesus is the "model of the godly life" who "has given his life for the benefit of others."[342] Jesus' life and work point the way to a new form of existence within the divine purposes of creation, and humans do well to model our own lives after his. Religiously and ethically, then, "Jesus becomes the central event for understanding what it means for humans to be God's

[339] Ibid.
[340] Ibid., 240.
[341] Ibid., 250.
[342] Ibid., 243-244.

proposal for the future of the evolutionary process."[343] The figure of Jesus as the redeemer is intended to show a life in accordance with the purposes of God for freedom, and in this, the life, ministry, and death of Jesus are intrinsically valuable. Jesus is not a supernatural being, nor God incarnate in any mystical or metaphysical sense, and so Hefner believes that these aspects of traditional Christian theology must be "rendered" and reinterpreted to fit a world of scientific naturalism, but Jesus still has a primary place in the lives of Christians as the one who not only showed but lived God's purpose in the creation.

The idea of humans being made in the image of God fits more easily in Hefner's theological anthropology, in that our co-creatorship analogously reflects the nature of the creator. Hefner argues that "to be created in the image of God implies that humans can be the vehicle for grace toward the creation, in a way that is somehow reminiscent of God's graciousness."[344] He analyzes some of the very different ways that the image of God has been interpreted throughout Christian history and the scant biblical warrants for any one of them as a dominant metaphor. Understandably, he ends up identifying the image of God with the creative aspect of human persons, but not just creativity in general. For Hefner, "the image of God should be interpreted, consequently, in terms of what is the quintessence of human nature, from the perspective of how that nature may be said to be analogous to God. I suggest that what is at the core of this analogy today is the character of *Homo sapiens* as free creator of meanings, one who takes action based on those

[343] Ibid., 248.
[344] Ibid., 238.

meanings, and is also responsible for those meanings and actions."[345] These acts of the creation of meanings in action and responsibility for both the meanings and the actions are the primary ways that we image God and are the final word about human life. What we create, what we do, and what we are responsible for matter quite ultimately. Jesus illuminates the divine purposes of freedom and is the model of human action for us to follow, but the essence of Christian faith, for Hefner, is not the grace embodied in forgiveness or a substitutionary atonement of Christ. He believes that for the created co-creator, responsibility is the necessary stance, and that we should not look to the forgiveness of God to save us, as a paradigm of divine forgiveness "implies that the works of the co-creator are unimportant, unnecessary, or expendable."[346] For his created co-creators, "the reality of redemption is the fact that the artifacts of our co-creating are acceptable and are in fact accepted" by God in fulfilling God's purpose for the creation.[347] Humans have been given grace by God to be creative in our action and responsible for those actions, and it is in acting that we find our fulfillment and salvation.

Philip Hefner reboots, revises, and renders the Christian tradition of theological anthropology and begins with an evolutionary account of our origins, finding that the best understanding of human nature is in the theory of the created co-creator. This created co-creator is a free and responsible creature whose purpose, given by God through the structures and processes of evolution, is to be creative in order to birth freedom for the whole of creation. Major categories of theological anthropology, such as sinfulness,

[345] Ibid., 239.
[346] Ibid., 249.
[347] Ibid.

finitude, or the image of God, are reinterpreted in this light and given new meaning within the co-creator paradigm. This approach has gained a great deal of traction among theologians of science, some technology ethicists, and Christians who are generally supportive of advanced technological development, but it also has significant detractors. These detractors include ethicist Brent Waters, whose more traditional theological anthropology also starts with the same evolutionary beginnings but leads in a very different direction. It is to his theological anthropology that I will now turn.

Brent Waters and the Embodied-Finite-Mortal Creature

Theological ethicist Brent Waters rejects and criticizes the created co-creator model of theological anthropology put forward by Philip Hefner and instead puts forward his own understanding of what it means to be a human person, particularly in light of contemporary issues in emerging technology that propose to change humanity's nature. His theological anthropology is intended to be universalistic, like Hefner's, applying to all human beings in all places and at all times. But the particularities he develops within his substantive understanding are written to speak to the situations of contemporary humans, particularly in the West, whose lives are those of "late modern…high-tech nomads."[348] He begins, like Hefner, with the understanding that humans have been formed in particular ways by our evolutionary heritage, and the particularities of that development have given us certain tools that humans (relatively) unique among the animals. For instance, "the ability to trust and cooperate with strangers is a rare phenomenon. Human beings are

[348] Waters, *This Mortal Flesh*, 19.

evidently the only species that has developed this capacity…it seemingly runs counter to the nature that evolution has bequeathed to humans."[349] Our biological heritage has allowed us to develop culture - knowledge and practices that are passed down across generations that are not directly based in our genes - and over the course of human history, this culture and its suprabiological process of evolution has come to be "the dominant, formative feature of human life."[350]

The cultural evolution of human beings, running alongside and now outpacing biological evolution, has produced technology as a key cultural feature, and under "late modernity," technology has become the dominant feature of culture that now defines our lives. Indeed, following Heidegger, Waters claims that "technology is the ontology of late modernity," meaning that our current culture "places its hope in the technological mastery of nature and human nature."[351] Given this set of assumptions about the nature of current culture and value in Western life, Waters' anthropological questions take on both ontological and ethical forms. For Waters, "to think about technology is to ponder the question of our very being as late moderns: who are we, and what are we hoping to become?"[352] What is at stake for him is both the meaning and preservation of humanness within a technological context that threatens both human existence and meaningfulness, and it is from this that his primary ethical questions arise. "We cannot define who we are or express what we aspire to become in the absence of technology…What are the

[349] Ibid., 23.
[350] Ibid.
[351] Ibid., 15.
[352] Ibid., 17.

underlying value and convictions that are operative in placing our confidence in technologies that purportedly enable us to live better lives?"[353]

Waters is concerned that we are becoming overreliant upon our technologies and that "the growing formative influence of technology takes on an aura of inevitability, for humans grow steadily more dependent upon the instruments they employ in their mastery of nature."[354] His most grave ethical and anthropological concern, though is that our biological evolution and our cultural production have become mismatched such that we begin to feel that we need to change our underlying biology to keep up with culture. Because our biological evolution has left us unprepared for our current cultural situation and moves far more slowly than cultural evolution, Waters argues that humans are becoming pressured to "transform themselves into beings that are fully formed by their cultural evolution; they must become more fully the product of their own purposeful will than outcomes of natural selection."[355] This pressure to direct our own evolution is far from benign, in Waters' thought, for it has the potential to rob us of our humanity. He fears that "to live well in the world that humans are creating for themselves, they also need to transform themselves into something better than human, a superior posthuman being."[356] It is in response to the threat of humanity giving up its own identity, meaning, and existence in order to become something "better" that Waters develops his theological

---

[353] Ibid.
[354] Ibid., 24. Interestingly, while Waters does not exempt himself from late modernity and its use of technology, when he writes about the problems of technology in late modernity, he refers to humanity as "they," instead of "we." But when he refers to the human condition theologically, the voice of his prose shifts and humanity becomes "we" again.
[355] Ibid., 24.
[356] Ibid., 27.

anthropology, to maintain our being as humans (albeit humans in a technological society).

He develops his anthropology using a disparate group of theological and philosophical sources. He derives his diagnosis of late modern technological society and its ontology from Heidegger's existentialism, and he uses Hannah Arendt's understanding of natality and mortality as definitive of the human condition. The relationship between necessity and goodness in relation to theological anthropology come from Simone Weil through the work of George Grant. Most importantly, Waters' theological anthropology borrows heavily from his mentor, Oliver O'Donovan, particularly in his distinction between the natural and the artificial and concerns about the degradation of humanity within the self-madeness of technoculture.

Waters' theological method is, in many ways, more traditional than Hefner's, in that he relies more heavily on classical Christian doctrines and affirmations, though he does not discount contemporary scientific knowledge in framing those doctrines. Waters examines theological anthropology's potential contributions to human self-understanding within contemporary technoculture's tendencies for remaking and self-improvement, remarking that "there are no free benefits; whatever is taken entails an unavoidable cost."[357] Within this, he seeks to explore what some of the costs of our technological remaking of the human are and how theology can be used to respond to these costs in order to maintain a fully human existence.

His method begins with two key theological insights from which he develops his anthropology. First, he argues that in Christian theology, "anthropology *is*

[357] Ibid., 31.

Christology…we turn to Christ to learn what being human means, and to catch a glimpse of our destiny as a species."[358] While his claim is somewhat stark on its face, this understanding follows the classical claims about how we know what it means to be a human being coming from understanding and being transformed by the life and work of the incarnate Christ. Waters' claims here are not Trinitarian in nature; he has no real interest in "personhood" or the potentiality of nonhuman persons, as his concerns are specifically about the human remaining human amidst technology, but his understanding of the human being can still be useful in understanding certain features of human existence in developing a broader theological anthropology.

His second key methodological claim follows O'Donovan's work on a "natural ethic" in which "one may choose to either conform to or rebel against an objective moral order. 'The way the universe *is* determines how man *ought* to behave himself in it.'"[359] In this line of reasoning, the created order of the world has been vindicated and upheld by God in the incarnation and resurrection of Christ, so we know in this that the created order is, in fact, good, as given by God. Just as we come to know what it means to be human in the incarnation of Jesus Christ, he believes that we also come to know the good in creation, as "it is in and through Christ that nature, which Christians properly name 'creation,' discloses its vindicated order that can then be discerned."[360] Through the resurrection of Christ, in the proper orders of creation, we then come to know the goals and meaning of human life, as well as our proper hope for genuine transformation. In the

[358] Ibid., 159.
[359] Ibid., 126.
[360] Ibid.

resurrection of Christ, God "discloses a *created order* that provides an objective standard and teleological order," which "provides the foundation of *obedient freedom*."[361] Rather than putting our faith for the future of humanity in technology, Waters exhorts his readers that "there is a proleptic trajectory revealed in the resurrection of the incarnate one, disclosing creation's destiny in the exalted Christ and his parousia. Consequently, the resulting ethic in teleological, rather than restorationist."[362] So, while Waters grounds his thought in a form of natural law, it is not simply a natural law that we can automatically read off of nature alone; it requires an understanding of the purposes of the creation revealed in Christ, though it is an "objective" natural law and order for the world that is shown and can be discerned.

The core of his proposal, then, is to analyze "what it means to take mortal and finite bodies seriously, since they have been affirmed, vindicated, and redeemed by God in Christ, the Word made flesh, particularly in light of current attempts to overcome the limits of finitude and mortality."[363] Waters believes that the vindication of creation by God in Christ points to a particular character of life in the flesh. First, life is fundamentally a gift of God, and thus to be appreciated and cared for as such a valuable gift. Second, the character of that life is one that is embodied, finite, and mortal; these conditions are not problems to be solved, but core characteristics of the gift of human life to be affirmed and celebrated in the way we live our lives. Third, we humans come with innate, insurmountable, and ultimately necessary and good limits. And finally, true

[361] Ibid., 159, emphasis original.
[362] Ibid., 125.
[363] Ibid., 9.

freedom is not found in attempting to abolish those limits, but to live within them. Each of these points will be addressed in turn.

First, Waters emphasizes the gifted character of human life, and its essential quality as that which is both our own and not our own. We are stewards of the gift and are intended to appreciate and accept the particular gift that we have been given. For "life is not a product that we produce or own, but a gift that is entrusted to us, and we are to care for and use this gift in accordance with God's expectations and commands."[364] The gift of life implies that we are entrusted to care for that life, but not to change it or to make it something other than the gift that has been given. The giftedness of life puts limits on us in changing who and what we are, both innate limits and moral limits. Waters does not entirely reject the use of technology or medicine to alleviate human suffering, but these practices are both bound by the gifted character of human existence under God.

Second, the gift of human life is characterized by three basic facts: "Humans are not only creatures; they are embodied creatures. As such they are also finite and mortal beings, and therefore subject to bodily limitations."[365] Our embodiment as living creatures in a physical / material world necessitates both our finitude and our mortality. In our being physical creatures, we are subject to the limits of physics, chemistry, and biology in both our being and our action. We are finite in our power and reach, as well as in our capacity to determine and change who we are. Our finitude arises largely from what roboticists would understand as situatedness, the historical, cultural, and material particularities in

[364] Ibid., 143.
[365] Ibid., 157.

which we live, but also from the fact that we cannot enact all that we may want. We are also, necessarily and blessedly mortal, for the biology of this world dictates that one generation must pass to the next, and evolution requires mortality in order to adapt, survive, and thrive. These three characteristics are intended to show both the goodness and the limitation of life, and that the limitations of life are themselves good, and are not to be overcome with our technology. Despite our attempts at self-transformation, we are bound by our embodied finitude and mortality, for "despite the progressive mastery of human nature, stubborn natural and biological limits remain."[366]

Waters and Hefner agree soundly, though for somewhat different reasons, on the necessity of genuine materiality of human embodiment. Waters seeks to defend a theological anthropology based on human embodiment from problems on two different fronts; first, from other Christians who embrace a body / soul dualism, and second, from the technological imperative to change the body to become something other than human. Throughout his work Waters has insisted on a Christian doctrine of humanity that is non-dualistic, as "a dualistic understanding of persons…is incompatible with our status as God's creatures. Humans are not composed of a body *and* a soul. Rather, God has created us *as* embodied souls and ensouled bodies."[367] Our nature as embodied creatures is a matter of faithfulness to the gospel of Jesus, for it was as an embodied human being that Christ lived, died, and was resurrected; no dualistic understanding of human persons can truly make sense of this. Waters argues that it was with good reason that the early church

[366] Ibid., 24.
[367] Waters, *Reproductive Technology*, 3.

condemned as heretical Gnostic forms of dualism, in which the body was a material prison for the immortal and all-important soul.[368] Christians, properly speaking, understand that

> Our lives belong to God, and it is in and through our bodies that God entrusts to us the gift and loan of life. Even attaching the qualifying 'embodied' to our status as 'creatures' is inadequate if it implies that ultimately the body is just a temporary container for the more important soul. It is not just the soul (or mind or will) that God loves and redeems, but the whole, full, and complete creature that bears and embodies the divine image and likeness. In this respect, it is important to remember that for many of our ancestors, it was the resurrection of the body, and *not* the survival of the soul, that seized their attention.[369]

It is not simply in, but as, bodies that we live our lives and engage the world around us, and it is as bodies that we have our hope in an incarnate and resurrected redeemer. But Waters is also concerned about the problem of dualism arising from technological attempts to master and subdue the body in order to transform it into something else. Under late modernity, we experience both a fascination with and a loathing of the body, in its vulnerability and limitations. It is against the views of thinkers like Kurzweil and Moravec that Waters makes his claims about both the necessity and goodness of the human body, for Christian "hope does not reside in transforming flesh into data, but in the Word made flesh."[370]

Our nature as embodied creatures likewise entails both finitude and mortality as inherent conditions. We strive to overcome our limits through technology, seeking to extend life toward immortality and an infinite grasp. But Waters argues that "in the absence of definitive limits, the resulting frenetic activity is enslaving…in contrast,

---

[368] Waters, *This Mortal Flesh*, 32.
[369] Ibid., 56.
[370] Ibid., 32.

affirming the finitude and mortality of our creaturely status provides an underlying order that is the prerequisite of freedom."[371] That is, when we try to constantly improve upon ourselves and our biology, we are doomed to being dependent upon that technology and the imperative to keep moving toward an unattainable and indefinite goal. We can instead choose the path of acceptance of our biological limits and mortality. Our status as finite, mortal creatures is not inherently problematic, evil, or even tragic. He says that instead, "evil and tragedy occur when humans, as finite and mortal creatures, attempt to vanquish death on their own terms."[372] Waters references the horrors of the twentieth-century experiments in totalitarianism as reason to believe that the yearning for technological mastery of the body is both nihilistic and threatening to freedom and human meaning. Technological striving to overcome the limits of the body constitute "an ill-advised attempt to strip away the vulnerability and imperfections that enable humans to be human and humane. It is not the death of humankind, but its humanity, that is at stake."[373]

We are creatures bound by limits, and these limits do not impinge on who we are or what God intended us to be, but are absolutely necessary to our being human creatures of God. "It is in their fragility and vulnerability, as well as the interdependence accompanying these qualities, that makes humans genuinely human. In their absence humans would be unable to learn the trust that makes their enriched social lives possible. Or in theological terms, it is as embodied creatures that they bear the *imago Dei* - the

[371] Ibid., 39.
[372] Ibid., 71.
[373] Ibid., 152.

image of their triune creator."[374] It is precisely in, and not despite, our embodied finitude and mortality that we become the humans that we are intended by God to be. Our finitude and mortality are not something imposed on us by the fall into sin, and our embodiment is not a trap from which we must try to escape. Rather, Waters argues that we are intended to realize the beauty of life as limited creatures. We must come to understand that our lives, "however vulnerable, fragile, and imperfect they might be, are nonetheless good precisely because they have been created and blessed by God."[375]

For Waters, our embodied, finite mortality does not take away our freedom, but makes it possible. "True freedom is a gift of the Spirit that frees us to be obedient to the definitive limits that delineate our lives as finite and mortal creatures."[376] Though we late moderns tend to think of ourselves as free when we overcome limits in order to make choices about our lives, Waters believes that this is merely an illusion of freedom, for each of our technologies that we use in our choices is built with intended purposes that we cannot work around, but that instead shape us and our choices. When we think that we are using technology to make ourselves more free, we simply become more deeply dependent upon our technology and the limitations on important choices that technology entails.[377] We cannot be free by trying to eliminate all possibility of unhappiness or suffering, for that can only be done by eliminating goodness and love. Our freedom comes by living and loving within the limits we have been given, and our doing so is blessed by

[374] Ibid., 32.
[375] Ibid., 159-160.
[376] Ibid., 127.
[377] Ibid., 42.

God. He argues that "to be genuinely free *and* human requires living dangerously, but not recklessly; that in and through love the risk of being unhappy is accepted."[378] In taking the risk of love within our human limitations, in obedience with the commands of God, we find true freedom.

We come to understand the intended character of human life and action through the person and work of Jesus Christ, in whose "incarnation the necessity of finitude and mortality, of human limitations more broadly, is affirmed rather than condemned."[379] The incarnation of Christ provides more than an example of human moral life for us to follow. It is the eternal God reaching out to the creation in love, redeeming us in, and not from, our finitude. For Christ came into the world as a human being in order to redeem human beings and all of creation. The theology of the incarnation indicates that God is not only creative, but loving, ever extending love to the creation. Waters argues,

> If the incarnation is to have any significant import…they must begin with the insistence that the finitude entailed in our status as embodied creatures is not merely an unfortunate limit to be overcome but defines and delineates the normative shape and pattern of human life within the dictates of temporal necessity. Why else would the Word who became flesh share in the most common experiences of birth and death? And why else the need for a crucified reconciler, resurrected savior, and exalted lord as the instrument of God's vindication and redemption? If flesh is ultimately not important, then Gnosticism, in both its late modern and postmodern manifestations, offers a far more appealing salvific story: escape, flee, or otherwise overcome any and all biological limitations.[380]

The story of human life that culminates in the incarnation of Jesus is one in which we are transformed, but it is not a transformation of our own making. Rather, it is one in which

[378] Ibid., 38.
[379] Ibid., 59.
[380] Ibid., 129.

the power of the Spirit reshapes and remolds us into the image of Christ. The destiny of the world is inaugurated and given its direction by Christ.

We may then ask what Waters believes that the relationship should be between Christians and the technological world we inhabit. For if we are defined as humans by our limits, accepting our fate in faith and awaiting transformation in Christ, then it seems that technology (including medical technology) may have no appropriate place in Christian life and practice. And Waters is indeed highly skeptical regarding the effects of many existing and emerging technologies on human culture and faith. His diagnosis is that late modern society is inhabited by "creatures who construct a reality that they will into being through the power of their technology," seeking "mastery of nature and human nature."[381] For Waters, this will to power administered through technology and the hope that such technologies will save us by freeing us from our limitations is slowly destroying us from the inside. He is troubled that the development of "therapeutic, preventive, and enhancement technologies is part of the larger task of blurring the line separating the natural and the artificial," and that as we become increasingly products of our own making we become "reduced to underlying information that can be manipulated."[382]

Yet Waters is not entirely anti-technology. He is concerned with our dependence upon it for self-transformation and the idolatry of hoping that technology will save us, but he does not reject technology simply because it is technology. Although with technology

[381] Ibid., 18, 15.
[382] Ibid., 25, 182. His fear is that an anthropology of information, which is precisely that which is suggested by several of the robotic futurists, is what unmakes our humanity and humaneness.

comes the possibility of peril, some technologies can genuinely be used for the improvement of human life, and this translates into a matter of discernment within the church's practices of justice and care. For "although the church's pastoral ministry should help individuals accept the (often unwanted) limits of their lives as embodied creatures, this is not synonymous with stoic resignation or silent acquiescence to fate."[383] The church does properly have a ministry of healing and an acceptance of medical care, including at least some forms of technologized medical care. This is because although we are finite and mortal creatures whose lives necessarily entail sorrow and suffering, "pain, suffering, and unhappiness are *not* goods to be pursued."[384] Likewise, while we are not to pursue immortality in violation of our limits as mortal creatures, we continue to recognize that death is the final enemy, albeit one that has (proleptically) been conquered in Christ. Waters argues that "there is no need to deny that modern technology and medicine have improved the lives of many people, and there is no compelling reason to attempt a nostalgic return to a more primitive age."[385]

Hefner and Waters Reframed and Reformed

Both Hefner and Waters insist on developing universal theological anthropologies;

[383] Ibid., 54.
[384] Ibid., 37.
[385] Ibid., 38. In making this claim, he is cautioning some fellow Christians who have tended to overstate the problems with emerging medical technologies and moved toward their rejection through nostalgia for a better world without technology. Waters recognizes that technology, including sanitation, antibiotics, and vaccinations, have vastly improved both the health and longevity of human beings and have prevented a great deal of suffering.

their attention is to humanity as a group of individuals that share fundamental characteristics. Hefner even goes so far as to argue that his theory of the created co-creator summarizes everything that one should know about a human being, while Waters places his primary attention on the universal limitations and vulnerabilities of the human condition. And this universalism is helpful in certain respects. It highlights basic similarities among humans that points to an underlying equality of moral valuation, rather than pointing to differences that may lead to some being valued more highly than others, or contributing to historic discrimination against certain people based on particular capacities (or the lack thereof). But in this attention to universally-shared features of human experience, they fail to attend to the real differences in experience that make us the particular people that we are. They extrapolate from their own experiences, both freeing and constraining, that are features of their being educated, straight, white, Western men, and assume that whatever they say will be universally applicable and recognizable to all. For instance, Hefner's insistence that the essence of human life is creativity arising out of freedom must be understood in his overall social, political, and economic context, for he has never known himself to be anything other than creative and free. Even as he grants space for human conditionedness by evolutionary forces, his understanding of human persons envisions little current external or internal constraint on freedom.

Both Waters and Hefner place a strong emphasis on the embodied nature of human life, though they do so with somewhat different emphases. Hefner has little to say about embodied human life per se, but he emphasizes instead the evolutionary heritage of humanity as fully natural creatures, conditioned by our environment to become free. He

rejects supernaturalism and gives no hint of dualism in his work; for him, human beings are strictly material beings, albeit beings with a spiritual connection to God (or the ultimate) within that materiality. But his focus on the essential freedom of humanity indicates little place for the reality of both physical and social constraints on human freedom beyond evolutionary and cultural conditioning. The particularities and messiness of embodiment and emotion are not factored into his thought on human nature. There is no strong sense of human beings as ever choosing to do evil, hurtful things to one another, but simply our fallibility in trying to live out our lives as best we can. He believes that this can lead to real evil, but the connection between the particularities of embodiment, the differences of experience in taking constraint seriously, and human tendencies toward greed and violence are not given any strong place in his work. This can lead to one form of romanticization of humanity: in our freedom we sometimes make mistakes that can cause real harm, but we do not have any innate problems in our embodied lives beyond finitude of knowledge, and thus the importance of embodiment and the likelihood of evil are understated.

Waters tends to romanticize embodiment in an entirely different direction, celebrating rather than downplaying finitude and joyfully embracing fate rather than working to overcome injustice. Waters is careful to highlight the absolutely embodied nature of human beings, defending against both Christian and posthumanist tendencies toward forms of modern-day Gnosticism that would separate the life of the body from the life of the soul or mind. In this, he repeatedly focuses on the holism of Christian theologies of embodiment, noting that any attempts to denigrate or escape the body in

favor of the soul have been soundly rejected throughout the history of the tradition.[386] For him, Christians do not hope for the survival of the soul upon the death of the body, nor for a technological revolution for biological or informational immortality, but for the resurrection of the body inaugurated and delivered by an incarnate God, who lived and died as an embodied, mortal human being. This emphasis on the embodied life of humanity is helpful in recognizing the messiness of human life, including his rejection of attempts to escape our human condition through various means. But he ends up claiming that finitude and mortality are not simply facts of human life that cannot be escaped, but are to be celebrated as the ground of our freedom. His focus on acceptance of finitude and mortality leads to a romanticizing of the pain and suffering of life as necessary, if not desired, and his emphasis on the acceptance of fate and necessity in bodily life move toward both a rejection of technology (including medical technology) and a political quietism favoring the acceptance of suffering.[387] I believe that it is important to resist the temptation to romanticize the body and its failings, but rather to recognize the body's importance, specifically in its messiness and suffering that make medical care and healing important Christian callings.

We can now turn to some ways in which they can each be used as lenses through

[386] Ibid., 56, 129, 158.
[387] Waters does indicate that he does not think that suffering is inherently good or to be celebrated, nor does he reject technology entirely, but he does not give much indication of why either of these things should be so, and they fit poorly into his overall system of claims about human nature.

which to view theological anthropology in a new frame.[388] They are not fully reconcilable, to be sure, but they provide poles of thought in which to situate Christian reflection on personhood, including, but not limited to the human form of personhood. Using the work of Hefner and Waters, I believe that the picture that emerges is one of a theologically-grounded form of embodied, sociable / relational, and concretely situated human life and personhood. While neither of them addresses the question of persons beyond the human, I want to argue that it is possible to use their theological anthropology to open up space for the possibility of persons beyond the human, which will be explored in greater depth in chapter 5.

Both theologians begin with an evolutionary account of human life and human origins, framed in the context of the world as God's good creation out of nothing. They likewise trace very similar narratives of the relationship between biological evolution and cultural evolution, arguing that our biological heritage has given us brains that we have used to develop culture, namely behavior transmitted across generations through learning rather than genetics.[389] Our cultural evolution - learned behaviors like language, technology, and religion - have now come to be the primary force driving changes in human life and form our situatedness in biocultural evolutionary history. Where biological evolution moves very slowly, over the course of eons, cultural evolution moves very rapidly, and humans have the adaptability to incorporate a wide variety of cultural changes into our everyday lives. We continue to be shaped, conditioned, and constrained

[388] Both theologians also utilize very problematic forms of natural law that are beyond the scope of this inquiry.
[389] Waters, *This Mortal Flesh*, 23. Hefner, *The Human Factor*, 19, 28.

by our biology, but the evolution of culture has given us a mode in which to become responsible actors and is now "the dominant, formative feature of human life."[390]

This grounding of their respective anthropologies in human beings' evolutionary heritage gives rise to a serious consideration of human beings' nature as embodied, natural creatures. Here, Hefner argues forcefully that we are entirely natural, material beings who were created to relate to God and one another through our evolved, biocultural embodiment. We have a nature that is given to us by God through the evolutionary process, and that we need no recourse to a non-material or supernatural understanding of human beings in order to properly relate to the God who made us. That is to say, it is in our natural physicality, our evolved biological life, and not through any non-material element, that we live in relation to God.[391] Waters argues similarly that our embodiment is the first and most important understanding of our lives as created persons. For Waters, our evolutionary heritage has given us lives as bodies that are fragile, finite, and mortal. We are not only fully physical, we are also "limited and highly dependent creatures."[392] Likewise, it is in and through the body that the incarnation of God in Christ is salvific for humanity and all creation. For unless our physical embodiment, in all of its joy and pain, was the essence of our being as human persons, there would be no reason for the Word of God to take on human fleshly life, to live and die as an embodied human person, and to be raised again as a body.[393] Neither Hefner nor Waters envisions a model of human life that

[390] Waters, *This Mortal Flesh*, 23.
[391] Hefner, *The Human Factor*, 19.
[392] Waters, *This Mortal Flesh*, 31.
[393] Ibid., 56.

is anything other than holistically embodied in its spirituality. Waters particularly refutes any Christian conception of created human life that denigrates the life of the body or finds the essence of humanity in an immaterial soul or mind instead of in the body.

Each emphasizes different facets of this embodied life, though. For Hefner, our evolved biological life is one in which freedom has arisen as an emergent property of humanity through the plasticity and adaptability of our bodies and brains.[394] The freedom exists within the context of our created, conditioned biology, but it is true freedom and is the essence of who we are as persons made in the image of God. Our freedom, for Hefner, is not simply the freedom to do whatever we wish, though; with freedom comes great responsibility for what we do with that freedom.[395] It is in our embodiment, through our freedom, that we relate to God by living out God's good purposes in and for the world. When we take responsibility for our actions as created co-creators to promote wholeness and flourishing of one another and of all creation, Hefner argues that we fulfill the good purposes of God. Our embodied spirituality comes with great freedom and responsibility, as we see in Jesus the perfect human life of giving oneself for others.[396]

For Waters, our embodied lives are also the ground of our spirituality, though this takes on a very different shape. In his thought, "humans consent to their mortal and finite limits because they are creatures who have been created in the image and likeness of God, and it is as embodied creatures that they love, serve, and are in fellowship with God."[397]

---

[394] Hefner, *The Human Factor*, 30.
[395] Ibid., 30-31.
[396] Ibid., 243-244.
[397] Waters, *This Mortal Flesh*, 158, emphasis removed.

Waters argues that our existence as spiritual creatures is not separate from, much less antithetical to, our embodied life, but rather arises within it; our lives as finite, embodied creatures are given as good gifts from God and are not to be despised.[398] It is within our lives as embodied creatures that we relate to God and other creatures, both theologically and ethically. We are situated in our circumstances in time and culture, elected by God to be transformed into the image of the risen Christ.

All is not well, however, in either Waters' or Hefner's accounts of embodied human life. We have been created good by God, but sin is a reality in the world that corrupts creaturely existence, causing suffering and hardship. Waters sees the greatest temptation to sin in contemporary technologized life in the attempt to overcome human limits, to become something other than finite and mortal, to master both nature and human nature.[399] He believes that our finitude and mortality are good creations of God, and that we are not to become something else, even in our desire to heal the sick bodies of creatures. Our sociable, relational lives are intended to be focused on caring and acceptance of the situations in which we live. We are not to seek to overcome our limits, but rather to accept our fate together. Therein lies too much potential for manipulation of our humanity, a rejection of the gifts of God, and a reduction of human life to the will to power.[400] Our lives, as the embodied finite persons that we are, are not inherently tragic, for Waters. It is not in recognizing and accepting our limits that we go astray. But we are also not to find goodness in suffering, so we need not reject technological forms of healing

[398] Ibid., 129, 158.
[399] Ibid., 28-29, 31.
[400] Ibid., 29, 71, 156.

entirely.[401] Our development and use of technology, though, should not lead to a reliance on technology instead of God for our salvation. In our late modern times, we are most likely to sin, in Waters' account, when we "attempt to vanquish death on [our] own terms."[402]

Hefner's account of sin, in contrast, is grounded in our limitations and finitude, not in an attempt to overcome those limitations. Because we are finite, fallible creatures, we err even when we try to do right, and in the trial-and-error process of evolution (including ethical evolution) we are bound to fail sometimes and to fall into sinfulness.[403] Our concrete situatedness and the limitations that arise from our embodied relationships are necessary, but they also give rise to our sinfulness. Hefner's recent work bridges this evolutionary account of fallibility and sinfulness with Waters' account. Here, our rightly creative yearnings and work for the betterment of all creation fall into sin when we pretend that we are not the limited creatures that we are, when we view our creativity "in terms of triumphalism."[404] For "we are not God; we are finite and sinful…There is no human sense of the future that can somehow escape the flaws of finiteness and sinfulness."[405] In this view, we are rightly both created good and unavoidably sinful. We are also both appropriately creative, seeking to move beyond our current limitations and inherently limited, finite, and fallible. Both Hefner and Waters, in their own way, are

---

[401] Ibid., 38.
[402] Ibid., 71.
[403] Hefner, *The Human Factor*, 136.
[404] Philip Hefner, "The Animal that Aspires to be an Angel: The Challenge of Transhumanism," *Dialog: A Journal of Theology* 48, no. 2 (June 2009): 166.
[405] Ibid.

concerned about the selfishness of humanity in our use of technology. Waters argues for this in terms of a Nietzschean will to power, in which humans re-create ourselves and our world without a God-given vision of the good; this re-creation does not enable the flourishing of humanity or the rest of creation, but instead "becomes little more than a thin justification for narcissistic self-indulgence."[406] Hefner is similarly concerned that the goodness of our technological projects will be warped and twisted to selfish ends, arguing that "the danger is that we will be closed around ourselves, self-directed rather than other-directed; that we think that our knowledge and skill are to serve ourselves, rather than the well-being of the creation."[407] Thus, both argue that there is a rightful place of technology, but that we must always proceed with caution, as we have been truly been created good by God and given our brains to use for God's redemptive purposes in the world. In Hefner's words, we are "alienated from the original creation and are implicated in finitude, evil, and sin…sin permeates every aspect of human existence."[408]

As we live our lives as embodied creatures, marked by both the goodness of our creation and the sinfulness which inhabits every area of our lives, Hefner and Waters likewise both understand that our nature is marked by our finitude - our mortality, our limitations, our vulnerability, and our interdependence. They interpret this finitude quite differently, but both note it as a crucial, inherent part of the nature of created human persons. Waters notes our inherent, creaturely finitude and understands it as the ground of our freedom, for he argues that it is in recognizing the limitations that we have that we

[406] Waters, *This Mortal Flesh*, 156-157.
[407] Hefner, "The Animal that Aspires to be an Angel," 166.
[408] Ibid., 165.

find where we can properly act in accordance with the redemptive purposes of God. Our embracing of our finitude, Waters claims, allows us to properly order our desires, appreciating and accepting the giftedness of life.[409] In acknowledging our finitude, we can be grateful for the lives that we have instead of viewing our finitude as tragic.

Hefner likewise acknowledges our innate finitude but does not see it as the ground of our freedom, but rather its limitation. In Hefner's work, our recognition of our finitude (along with our sinfulness) should temper all of our pretensions as co-creators, for "the co-creator…is an inveterate sinner, a perpetuator of evil, even when best-intentioned; and frequently human intentions are far from good."[410] We must always be aware of our limitations, in Hefner's work, but these limitations are not the end of the story; indeed, they are not even the heart of who we are. In his understanding, "the co-creator must proceed in its awareness of finitude and sinfulness; and, like a competent engineer, factor that finitude and sinfulness into the imagination and into the activities of co-creating."[411] For Hefner there is an aspect of rightful, God-given transcendence that is also a part of who we are. It is this seeking to transcend the limitations with which we have been bestowed by evolution, to continue the creation for the purpose of flourishing, and to live out the purposes of God in the world that we find our freedom and become who we are intended to be. Our movement beyond ourselves is not, for Hefner, a usurping of the action of God; it is our proper action given by God, and is a point of grace.[412] Here,

---

[409] Waters, *This Mortal Flesh*, 45.
[410] Hefner, "The Animal that Aspires to be an Angel," 165.
[411] Ibid.
[412] Ibid.

Hefner's understanding of the place of the creature's transcendence of limits tempers Waters' emphasis on embracing our limits, just as both recognize our capacity and tendency to sin in our actions.

Both Waters and Hefner uphold the traditional Christian affirmation that human beings are made in the image of God, though they attach different content to this concept. For Hefner, we image God insofar as we are analogous to God, namely, in our creative activity as creatures made by God.[413] In this analogy, then, "to be created in the image of God implies that humans can be the vehicle toward the creation, in a way that is somehow reminiscent of God's graciousness."[414] Thus we are to image God through the transformation of the world through our creativity in the direction of God's gracious purposes for the flourishing of the whole creation. Waters' view of the image of God is more explicitly Christocentric, in that we image God when we image Jesus Christ. We are made in the image of God in our election to be the creatures who give order to the creation toward the divine purposes.[415] In this, Waters and Hefner agree that we are to be moral agents in the world, working toward the purposes of God. But Waters, in his Christocentric understanding, takes our imaging a step further. We are made in the image of God, and thus we are called to agency in our election to be in relationship with God, but we are also called to become the images of Christ, who himself is the image of God. In this, we are not to rely on our technologies to make us better, but instead to be

[413] Hefner, *The Human Factor*, 239.
[414] Ibid., 238.
[415] Waters, *This Mortal Flesh*, 181.

transformed in and through Christ, to become re-created in his image.[416]

I now want to draw on this reframed anthropology, in which the understandings of created human personhood developed by Hefner and Waters are used to balance and support each other, to develop a robust theological anthropology that can be used to understand the shape of human life, but also can be used to view the possibilities for created persons who are not human, for machines that could be understood as human-like persons in a real sense. Drawing on the anthropology of Hefner and Waters, I want to indicate how their theological reflections might add to the robotic futurist narrative of anthropology as emergent minds that are embodied, sociable, and situated in their enviroment. It is in the balancing of these aspects together that we come to see human persons in our richness and the possibilities for other created persons as well.

Embodiment, Sociality, and Situatedness in Christian Theological Anthropology

As humans are said to be made in the image of God, we relate to God as creatures made by God whom God has chosen for redemption. We live our lives as embodied creatures, and Christian thought has provided a wide range of reflections on the meaning of this embodiment and in what ways we as physical beings can be spiritual creatures as well. Here I want to pick up on the claims made by both Waters and Hefner regarding the necessity of understanding human beings first and foremost as embodied creatures and the importance of maintaining a holism in our understanding of created persons.

Early Christians within a Hellenized Jewish context sought to understand what,

[416] Ibid., 159.

exactly, it was that was redeemed by God in the work of Christ. On the one hand, Christ did not come and provide the political and religious liberation that was expected. His reign was not an earthly one, but a "heavenly" one.[417] His preaching routinely spoke of the "kingdom of heaven" and not primarily of a social or political transformation of earthly society. And yet, Jesus was intimately concerned with the care of the body for those to whom he preached. He fed the multitudes, healed the sick, and cast out demons that afflicted people in order for their bodies to have rest and peace in a weary world. He was born into a human, physical body, and he lived, suffered, and died as a human being, with all of the joys and sorrows of embodied human life. His reign was not of this world, but his disciples were, and Jesus both cared for their bodies and insisted that they care for the bodies of those around them who were in greatest need of care. The disciples were instructed to feed the hungry, clothe the naked, heal the sick, visit the prisoners, and care for the widows and orphans. The concerns of Jesus were manifold and concrete.

Some early Christians picked up on the spiritual ideas of their time and embraced wholeheartedly the idea that Christ's mission was to free the soul from the prison of the body and to bring the soul back to God. All of the extreme versions of this Gnosticism were roundly rejected by the early church because they failed to take seriously the material nature of Christ and Christ's mission.[418] As argued forcefully by Waters, contemporary theologies that emphasize the salvation of the soul to the neglect of the body fall into the same Gnostic trap and should be rejected, just as robotic anthropologies

[417] John 18:36.
[418] Waters, *This Mortal Flesh*, 56, 129, 158.

that emphasize only intelligence or patterns of the mind and reject the body should be rejected. But even with the rejection of forms of Christianity that saw the material world as fundamentally evil, the idea of human beings having a "soul," an essence of the person somehow distinguishable and separable from the body, persisted, despite the lack of such ideas in scripture. At its best, Christianity reflected on the salvation of the person as a whole being and the material creation as called "good" by God. In Christian theology, sin has indeed entered the world, but this is not a negation of the goodness of materiality or embodiment, only the occasion for its redemption.

It may be asked at this point what the place of the concept of a "soul" should be within Christian theology, and here I take a minority, but I believe important, position. With Hefner and Waters, I find that the concept of the soul has been one that distracts from the key doctrines of Christian theology and is in no way inherent to it. There are certainly those who have developed theologies that both take embodiment seriously and include "souls;" I do not wish to take away from them. My emphasis on embodiment within the context of Christian theology points to a neglect of the place of the body and materiality in both theology and dualistic futurism. With regard to the composition of the human being, I follow the work of Nancey Murphey that is also used by Hefner and Waters: "First, we are our bodies - there is no additional metaphysical element such as a mind or soul or spirit. But, second, this 'physicalist' position need not deny that we are intelligent, moral, and spiritual…we are *Spirited bodies*."[419] This avoids both forms of the

[419] Nancey Murphy, *Bodies and Souls, or Spirited Bodies?*, (New York: Cambridge University Press, 2006), ix, emphasis original.

dualistic anthropological problem - we are neither fully determined, mechanistic beings within our physicality nor are we really minds that are trapped in bodies that are in need of replacement, whether through a flight of the soul or transplantation into a more durable substrate. We are our bodies, and it is exactly as our bodies that we are our full, spiritual selves.

The key to understanding (doctrinally and ethically helpful) Christian reflection on the material world is in a consideration of the incarnation and resurrection of Christ. Human beings, along with the rest of the world of space-time, were created by God, called good, and continue to be sustained by God. Sin entered into this reality, and so also then it was this world into which the Logos of God came in order to redeem this world and restore it to God. It is for this reason that the resurrection takes such an important place in Christian theology. Christ lived and died as a fleshly, physical human being, in a particular time and place in history, just as humans live and die as fleshly, physical human beings in our own situatedness. When we die, our deaths are real deaths, not simply a separation of the body from the soul. If that were the case, Christians should look forward to death, even encourage it, so that the soul could be returned to God. But because we are physical beings, our deaths have a genuine finality to them; there is no soul that flies to God upon the death of the body. Because of the resurrection of Jesus, Christians have a hope in the coming general resurrection - the restoration to life of the person who died into a (perfected) body. Jesus was resurrected and returned to his followers. He was changed in ways they could not understand, yet he ate and drank with them and was recognized by his body. Even in its perfected, resurrected state, his body

bore the scars of that which he suffered physically in his life.[420]

All this is to say that the point (and scandal) of the incarnation (meat-ification) of God in the world was not to simply call people to repentance in order to "save souls." No fleshly redeemer is needed to save the disembodied soul from damnation. To that end, no physical creation is needed if the "soul" is what is to be saved. The creation of the universe was not, theologically speaking, just a big messy mistake on the way to spiritual glory, and so Christian spiritualities that over-emphasize place of the soul must be reconsidered in their entirety.

Rather, Christian theology properly presents the material body of the person and its historicity as the true person, and our spirituality arises from this physicality and not despite it. We relate to God as the bodies that we are, the bodies with which we have been created and called good. Just as we image God in our transformability into the likeness of Christ, so we participate in God through the material and the everyday of life. Jesus instituted the sacraments in order that people could access means of grace in ways that were both tangible and understandable to us - ordinary, physical objects like bread and water are the means by which we enjoy our spiritual lives with God. We do not participate in God more fully through some non-material means. We are called to ministries of the body in its fullness - it is as our bodies that we are both mind and spirit.

Hefner and Waters both also utilize aspects of both sociality / relationality and concrete situatedness in their theological anthropologies. They draw these concepts in

[420] This fact and its implications are commented on extensively by Nancy Eiesland in *The Disabled God: Toward a Liberatory Theology of Disability*, (Nashville: Abingdon Press, 1994).

somewhat different directions and apply slightly different language. This, I believe, provides a stronger picture of what both sociality and situatedness entail in created personal life. Waters focuses his anthropology on the finitude of the created human person. This finitude is primarily an issue of situatedness, for it is our rooting in concrete, historical, cultural, and physically limited space and time that we come to be who we are as persons. This situated finitude gives us a place from which to understand ourselves in the world and in our relationships as well as a space from which to act. Hefner instead emphasizes the freedom of humanity, albeit a situated freedom with a sociable / relational purpose. For Hefner, our freedom does not exist without a purpose; its purpose is our use of our co-creatorship to develop and sustain our relationships with God and one another. We are embedded in our evolutionary and cultural heritage and it is from this heritage that our freedom arises.

Just as Hefner's understanding of our creative transcendence can be used to temper Waters' sometimes excessive acceptance of finitude, so Waters' emphasis on limitations can be used to temper Hefner's understanding of the place of human freedom. For Hefner, freedom is the essence of human beings; we are free in order that we might bring freedom to all of creation.[421] With this freedom, that Hefner argues is the heart of theological anthropology, comes great responsibility - we created co-creators are God's agents in the creation, and we are accountable for what we do in and to that creation.[422] Waters also believes that we human persons are in a position of responsibility, but he

[421] Hefner, *The Human Factor*, 27.
[422] Ibid., 30-31.

argues against Hefner that the model of the created co-creators tries to take on too much, that the responsibility for creation that the model implies can be borne by God alone.[423] Both, however, find that in our embodiment and situatedness, our moral relationships with one another take shape under conditions of both freedom and finitude. For Waters, we do have real freedom, and this freedom comes with responsibility, but it is not unconstrained. Because of the limitations of our finitude and sinfulness, we are bound. We are vulnerable and interdependent and are not, properly speaking, free to so as we wish.

Feminist and womanist scholars have highlighted the importance of attending to the genuine differences in experience between human beings in our embodiment, relationships, and situations, particularly in the experience of constraint. Katie Cannon, for example, calls attention to the experiences of black women under slavery and their work of humanizing their lives and creating spaces of freedom when they were treated as chattel and given virtually no control over the basic qualities of their own lives. For "under slavery the Black woman had the status of property: her master had total power over her, and she and her children were denied the most elementary social bonds - family and kinship…Being both Black and female, the Black woman survived the most wanton misuse and abuse."[424] Cannon argues that these women under slavery fought mightily and creatively against their enslaved condition, but simply characterizing them as free to be creative is a disingenuous description. Theologians who write as though their experiences are universal, that the particularities of their situated embodiment do not matter, are

[423] Waters, *This Mortal Flesh*, 177.
[424] Katie Geneva Cannon, *Katie's Canon: Womanism and the Soul of the Black Community*, (New York: Continuum Publishing, 1995), 48-49.

misguided and necessarily end up with less-than-adequate accountings for the fullness of human life. A robust theological anthropology must attend to this difference, even while it attends to the universally-applicable aspects of human beings. Hefner partially acknowledges the importance of this situatedness in his understanding of the conditioning of humanity by evolution, and Waters strengthens the understanding with a recognition of our dependencies.

Here the reflections on embodiment and embeddedness given by both Waters and Hefner (in their distinctive forms) are helpful, though they each must be tempered as well. For Hefner, this means taking seriously the reality of human evolution, toward which he provides excellent contributions in claiming the ways in which our natural origins explain the particular conditionedness and freedom of human life. It should also be tempered by Waters' reflections on the finitude and mortality that human beings also inhabit. We do have freedom, and we are finite, but freedom and finitude are not all of who we are and must be considered in relationship to one another. We are embodied, sociable, and situated, and while that is critical in both robotic and theological anthropology, these facets are not the end of the story. We are both finite and transcendent, free and constrained, good and fallen, and it is in these paradoxes that we discover the richness of what it means to be a created person. Theological anthropologies that emphasize one side of the paradox over the other, such as freedom over constraint in Hefner's work and finitude over transcendence in Waters', provide a theologically and practically incomplete picture of human and other created persons.

Chapter 5 – Robotic Anthropology and the Future of Humanoid Robots

This final chapter will build on the theological considerations of personhood in chapter 4 and will seek to explore some ways in which these theological understandings might be applied more concretely in the ethical development of humanoid robots. Having analyzed some of the trends in anthropology developed in robotic futurism and some of the implications for helpful ways to think about robotic personhood that arise from that, I have sought to indicate how a theological reading of personhood might build upon and expand the view of persons as embodied, sociable, and situated. I will first look at the character and status of these theological reflections in an inclusive and pluralistic context - what it might look like to "widen the circle of grace" for justice and the flourishing of both the humans and machines of the future. I will then explore some possible features and concepts that might be useful in considering humanoid robots as the *objects* of human moral action. After that, I will explore possibilities and concepts from within futurism and theology in considering humanoid robots as potential *subjects* of moral action. Finally, I will seek to draw these ideas together to look at some practical applications of these ideas as robots as both objects and subjects of moral action in the three key application areas outlined in chapter 1: labor replacement and human services, military and defense, and sex and companionship.

Widening the Circle of Grace

Any robust theological anthropology must give some account of the ethical and future-oriented choices that we make, both individually and collectively. Our yearning for

transcendence, goodness, and wholeness is one that provides a picture of a full human life - offering joy, freedom, and flourishing for all. It is this vision of integrity within ourselves, with one another, and before God that drives Christian ethical vision. Our finitude, fallenness, and constraint are not the end of who we are. A Christian vision of the reign of God is one rooted in grace and driven by it, epistemically and ethically.

This grace is the hallmark of Christian faith, ethics, and hope. Grace within Christian theology has been understood in very different ways within different strands of the tradition - as the perfection of nature or the restoration of a relationship. In this work I want to describe it as the creation of goodness where it does not exist. This begins in God's own action of creation. Where nothing existed before, God created something, and it was good. It is the restorative counterpart to sin understood as a loss of integrity - grace restores that which was lost and creates the goodness of integrity where it did not exist. It can perfect nature or restore a relationship but is not limited to these analogies. On the contrary, it is something wholly new, coming in from nowhere, a goodness breaking in without regard to merit. It is grace on which both Christian hope and Christian ethics are based.

A Christian theology of personhood begins, not with human beings, but with the God who is Three-in-One, the divine community of persons, perfectly united as the God of love and grace, who creates, sustains, and redeems the world. The divine drama is one in which God covenants with the creation, calling the people of Israel to Godself and through that people offering the blessing of grace to all the world. God calls all persons into relationship with Godself - a relationship intended for joy and freedom. God offers

Godself in covenant with humanity, and though humanity errs and falls away from God, God sends a redeemer, always reaching out to the people and offering grace. God creates all that is, and in grace, God creates goodness where there was a lack. In this, all creation, all truth, all joy, all love, all beauty is grace, and we come to know the good by experiencing grace in our lives. We who are persons come to know grace personally, though the grace of God extends to all creation, as the creation and sustaining of the world is itself a gracious act of God. In Christ we see the grace of God perfected - perfect love and perfect unity between human and divine.

The grace of God in Christ is given for all (created) persons and is intended for the flourishing of all - the enjoyment of God forever. But the story of the Christian community has not always been one in which this universal calling and offering of grace has actually been extended to all people, or to all people equally. The church has had its ups and downs on this point, ethically and theologically. At its worst, Christian churches have tended to restrict the offer (or benefits) of the grace of God to only some persons, typically males of status. But in its better moments, Christianity has served as a pioneer pointing to the way of grace for all - widening the circle of who is "in" to both receive and give grace. This theological and moral progress has been far from linear; it is not simply a long march of progress, but a winding road that sometimes doubles back upon itself, rejects those it has accepted, and finds reasons to reject persons whom God has called into grace. Even now, there are contradictory tendencies with Christian churches and movements about who to accept as true "persons" created in the image of God. The church's circle of grace has slowly been widened so that women, people with disabilities,

people whose skin is dark, poor people, and same-gender loving people, among others, know that they are embraced by God's grace. But the churches have not always recognized the full extent of this universality, so in practice, the circle of public acceptance had to be widened in order to include those who had been excluded and appreciate them as fully called members of the community.[425]

This widening of the circle and its attendant effects, such as diversification of leadership and transformation of the way that those who have been excluded are welcomed and treated, constitute a form of genuine moral progress. And while up to this point the expanding circle has been one of recognition of the full personhood of those who are already persons in the sight of God, the circle may soon need to be widened again to include types of persons who do not yet exist. I believe that there are four possible categories of theological persons who existence (should it come to pass) would need to be recognized by the church as "personal," and in need of a relationship to God and others in a way that is likewise personal. Three of these types of possible persons - human-machine cyborgs, human-animal hybrids, and genetically posthuman persons - stand beyond the scope of this inquiry. But the fourth type - the "artificial" person of the humanoid robot - could be developed in such a way as to be considered theologically personal, receiving the grace of God in a uniquely personal way, learning grace through its interactions with others, and offering grace to others as a person.

Christian theology acknowledges that there are persons who exist who are not

[425] I want to be careful in how I claim what has happened in the widening circle. It was not that the call of God's grace did not exist previously; God's grace is a universal offering.

human beings. The archetype and perfection of the personal is the personhood of God - three personal hypostases in one substantive ousia. This perfect personhood is exhibited by God in relation to the world, and through the personhood of Christ we learn the intended shape of human life. Even among non-divine persons, humans have not stood alone in Christian theology. The angels (and in some cases, demons) serve as an alternative form of personhood that de-centers the specifically human form of personality. As one need not be human to be a person, ethically and theologically, there is room in the Christian narrative for new categories of created persons who are also not human, possibly including robots who live "personal" existences, and this chapter will explore ways in which that personhood could take shape in interaction with human persons.

The theology out of which the possibility for robotic personhood arises is uniquely and unabashedly Christian in nature, though I believe that one need not claim Christianity in order to have this understanding of different forms of personhood be useful. That is, while the understanding of personhood as "personality" and created personhood as being characterized by different sets of features and paradoxes is not the most common (secular) way of understanding persons, it does give insight as to what might be considered a "who" outside the context of simple rationality or sentience. It provides a greater context from which to consider robots as potential objects and subjects of moral action, and it is to these that I will now turn.

Robots as Objects of Moral Action

I want to first explore the various ways in which something can be an object of

moral action and the types of moral value that underlie it. How we treat the things that we make and why we treat them the ways that we do, both empirically and normatively, will be key here. Within this, then, I will consider the ways in which certain types of persons or person-like beings are identified as objects of (human) moral action. The last section here will explore the relationship between robots as objects of moral action and the humans who act.

Moral Action and Moral Value

Next I want to consider some of the ways in which someone or something can be the object of moral action. Being an object of moral action runs along a continuum of more direct and more indirect objects, which partially correlate to the object's moral value, whether intrinsic, extrinsic, or both. The form of moral value will be examined first, followed by the directness and reasons for moral action.

A being can be understood to have moral value in a variety of ways, classically intrinsic, extrinsic, or both.[426] That is, a being can have value in itself, for an instrumental purpose, or both. Most of the artifacts that humans create or encounter have solely extrinsic value. A toaster is valuable because of what it can do in preparing bread; it does not have value for its own sake. Technological artifacts are generally thought to have only extrinsic value - they have no life, no feeling, no personality of their own, and so they do not have value in themselves. They are good insofar as they do what they are intended to

[426] For the purposes of this discussion, I will limit this analysis to the matter of concrete things/persons, rather than ideas or virtues.

do in serving human beings or our interests. In contrast, works of art or beauty can can be considered (in some ways) to have only intrinsic value.[427] A painting does not serve any function on its own - it does not do anything instrumental to sustain life, perform tasks, or serve normal human purposes. It is beautiful in and for itself. It may be enjoyed by human beings, but that enjoyment is derived from its intrinsic status as a thing of beauty.

Human beings, along with (at least some) other living beings, have both intrinsic and extrinsic value, though the balance between the intrinsic and the extrinsic varies in weight. For instance, my cat has both intrinsic and extrinsic value, but her intrinsic value has a relatively lower priority than my own. Her life is, in itself, a good thing, and should not be wasted unnecessarily. But she also has an extrinsic value - being a companion to her humans - that is also very important. Human beings, in comparison, have extrinsic value that is weighted lower in relation to our intrinsic value. We are, in Kantian terms, beings who are to be treated as ends and not solely as means.[428] We do have extrinsic value, as we are capable of being instrumentally useful in a wide variety of capacities. We can work, help one another, cultivate and prepare food, build shelter, write books, solve problems, care for the sick, etc. Our extrinsic value is important and should not be overlooked. But that extrinsic value is always weighted in relation to our intrinsic value,

---

[427] This is certainly an arguable point, as works of art can serve as inspirations, points of delight, historical references, or serve human purposes in other ways. They do not, however, have the kind of direct instrumental value that is true of technological artifacts, clothing, food, or shelter.

[428] Immanuel Kant, *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor, (New York: Cambridge University Press, 1997). Here I am using the form of the categorical imperative without using the same criteria of rationality as the basis for determining which beings are means and which are ends.

which, while not unlimited, is great, and so any extrinsic value that we have stands under the intrinsic. Our intrinsic value is such that our extrinsic value cannot ethically be used for our diminishment. Human beings, as persons of great intrinsic value, should be treated as beings who have intrinsic value - as valuable in ourselves. We should not be discarded or destroyed without an overwhelmingly good reason.

Robots are currently technological artifacts that (generally) have only extrinsic value.[429] They are good in what they can do for us, but they have no life, no feeling, no reason to exist on their own. But this could change, and it is important to distinguish criteria by which one might come to see a robot as having intrinsic value. At what point does a robot begin to have more than solely instrumental / extrinsic value? A humanoid robot might be considered to be like my cat - having genuine, but limited, intrinsic value such that its existence is an inherently good thing and not to be wasted unnecessarily, but having primarily extrinsic value. But there may also come a point at which the robot's existence takes on a greater intrinsic value, more like that of a human being. Some possibilities for determining this will be explored in the next section.

Robots can also be objects of moral action in more direct or more indirect ways. They can be objects of moral action indirectly as a result of a general principle, because of the possibility for harm / good to another being of extrinsic value, because of the possibility of harm to another being of intrinsic value, because of a definite harm to another being of extrinsic value, and because of a definite harm to another being of

[429] There may be robots that exist also as aesthetic works of art, and so may have intrinsic value as things of beauty, but this is a rarer case that can be distinguished from an average robot of which there are many identical copies that are useful.

intrinsic value. A robot could also be the object of direct moral action because of its own intrinsic value.

First, robots could be the indirect objects of moral action because of a general principle. They could be treated well or not harmed because of a general rule against wastefulness, for instance. In this sense, the robot is not being treated as an object of moral action because of anything about the robot itself, but rather because of the time, energy, and materials that someone put into designing, developing, producing, and purchasing the robot. I do not throw away my Roomba because it is foolish and wasteful to do so when it serves the purposes for which it was built. The robot could also not be harmed here because of a general principle not to harm anything unnecessarily. While I may take apart my robot in order to study it, understand it, upgrade it, or fix it, I do not smash it for fun because there is no reason to do so.

Second, robots could be indirect objects of moral action because of the possibility of harm (or good) to another being of extrinsic value. In this, a robot is treated ethically in order that some other good might be had that is also of instrumental value. We hold prohibitions against kicking a robotic dog so that people are more likely not to kick real, living dogs. This does not mean that the good will necessarily follow, but the ethical treatment is performed in order for the possibility to exist.

Third, a robot could be treated as an indirect object of moral action in order for there to be a possibility of good for a being of intrinsic value. I care for the (power, software, maintenance) needs of a robotic nurse so that that robot can assist in the care of my mother. The maintenance of the robot does not necessarily lead to be best care of my

mother, but it opens the possibility for her care to be improved by helping her get out of bed, make breakfast, etc.

The fourth and fifth reasons why a robot could be treated indirectly as an object of moral action follow from the second and third, only the possibility of bringing about a good changes into a definite and direct good for the other being. I take care of my Roomba so that it may be able to clean my house. A clean house is something that only has instrumental value, like the Roomba, but when the robot facilitates the good of the clean house, both are treated as objects of moral action (such as maintenance and charging). In the fifth case, a landmine deactivating robot is developed, used, and maintained in order to save the lives of intrinsically valuable people. The saving of lives is the entire purpose behind its existence, and it has a direct good for which it is used, so it becomes an indirect object of moral action.

The robot can also be a direct object of moral action because of its own intrinsic value. Where the robot's extrinsic value causes it to be the indirect object of moral action, its intrinsic value could allow it to be the object of direct moral action. To the extent that I believe the robot to have a goodness in its existence apart from the ways in which it is used, I come to treat it as a direct object of moral action. I talk to a humanoid robot so that it can learn; I treat it well because I believe that it would suffer if harmed; I provide it with opportunities to develop relationships in order to have a "healthy" existence. These are all ways in which its intrinsic moral value is recognized and responded to by a human; it means that I think of the robot in ways that are more or less personal.

Robots, Pets, and Chimps

A humanoid robot that is treated as a being with at least some intrinsic value and with direct moral action can be compared to a number of other types of beings who have varying degrees of balance between intrinsic and extrinsic moral value. I want to argue that they all share certain characteristics that have been understood as underlying personhood in robotic futurism - embodiment, sociality, and situatedness, and that these may form appropriate bases on which humans judge intrinsic moral value. They still vary, though, and so I want to offer some speculative possibilities for why this might be the case for beings that we a) tend to treat that being as a person, and b) should likely treat that being as a person.

I believe that a being that exhibits the qualities of living embodiment, sociality, and situatedness can and should be considered as a being with at least some degree of intrinsic value, as these characteristics underlie the possibility of all of the theological characteristics. A cat is clearly an embodied, living being, and she is also, in ethically significant ways, both sociable and situated in her environment, and so should be treated as having intrinsic value. My cat has become who she is through the relationships that she is a part of with the people she lives with and others who spend time with her. She is a very sociable animal, always wanting to be where the people are. This is not simply a product of her nature but of what she has become in a sociable environment with other humans and cats. She is unique in the degree and form of expression of this sociality, and she relates to others in her own way. She is likewise situated in her particular time, place, and relationships. Who she has become and what she offers in her relationships with

others are based in her situatedness. She both reflects and is reflected in the environment she inhabits, in relationships, in the physical space, and in the patterns of life of which she is a part.

But there are certain qualities that are common features of human life that she lacks that may explain some of the variance in how cats and humans are treated. She is finite and situated but expresses no empirically-identifiable sense of transcendence. She may be both good and fallen as a creature in a good but fallen world but does not have the theological or moral sense to make these characteristics her own. She does have some freedom and a great deal of constraint. She chooses when to be loving and snuggly and when to bite, based on her own feelings and states of consciousness at the time, which are only partially related to her present external circumstances. She is also constrained in this freedom. She cannot leave the house except to go to the veterinarian, so her lived environment is very small. Her behavior is significantly regulated by the humans with whom she lives. She does not gather her own food, provide her own shelter, or care for her own health, though she does do what she can to nurture those with whom she lives. In all of these things, her "personhood" can be understood as somewhat qualitatively different from that generally experienced by human beings, though this does not mean that she lacks any moral value of her own. We humans are still responsible for her well-being, for treating her humanely, and for allowing her to have the opportunity for a healthy, happy life, simply because she has intrinsic value.

A chimpanzee shares the same characteristics of embodiment, sociality, and situatedness, and so also has intrinsic value, though it also can exhibit some of the

theologically personal qualities that are frequently attributable to humans. Like my cat, a chimp does not (identifiably) yearn for the divine or some reality beyond himself; he does not seek a deeper meaning in the world around him. But a chimp does understand his place within a community and does care for those around him. Chimps, like elephants, exhibit grieving behaviors when someone they are close to dies.[430] They seem to understand, if not transcendence, at least loss, and the fact that death is a permanent separation. Chimps express distress at loss or conflict and provide reassurance and stress alleviation to one another.[431] They exhibit qualities of goodness and fallenness, as they can both lie and feel anger at being lied to.[432] They can plan aggressive acts toward one another; for instance, one will choose a large rock (going through a series of ever-larger rocks) in order to intimidate another at a later point in time.[433] They can learn and transmit that knowledge to the next generation, whether in tool-usage or food washing before consumption.[434] In this they express both the freedom to learn and develop new ways of interacting with their environment as well as the constraint of being taught. They choose particular relationships, mates, and food among a range of options. Like humans, they remember the

---

[430] Frans de Waal, *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*, (Cambridge, MA: Harvard University Press, 1996), 53-57.
[431] Flippo Aureli and Orlaith N. Fraser, "Distress Alleviation in Monkeys and Apes: Window into the Primate Mind?," *The Primate Mind: Built to Connect with Other Minds*, ed. Frans B. M. de Waal and Pier Francesco Ferrari, (Cambridge, MA: Harvard University Press, 2012), 246-264.
[432] de Waal, *Good Natured*, 76-77.
[433] Frans de Waal, *Peacemaking Among Primates*, (Cambridge, MA: Harvard University Press, 1989), 38-39.
[434] Toshisada Nishida, "Individuality and Flexibility of Cultural Behavior Patterns in Chimpanzees," *Animal Social Complexity: Intelligence, Culture, and Individualized Societies*, ed. Frans B. M. de Waal and Peter L. Tyack, (Cambridge, MA: Harvard University Press, 2003), 392-413.

actions of others and make future decisions based on those actions. Because of this, we tend to identify more with chimpanzees than cats and are more likely to recognize them as fellow "person-like" beings. We tend to ascribe to them a higher moral status and treat them as beings with interests and values of their own.

The personhood of a robot, then, can also be rooted in its intrinsic moral status as it is embodied, sociable, and situated. This does not, strictly speaking, depend on any non-mechanistic sense of "consciousness" in the mysterious sense of human self-awareness. But it might mean that for humans to understand and treat robots like beings with intrinsic moral value, they should exhibit certain qualities, and we are likely to ascribe higher intrinsic value to them based on the form and quantity of the qualities exhibited. Currently, robots are embodied, and there is research happening to make them more sociable and situated in their environment and relationships. They do not generally have the latter two qualities in most commercially-available applications, and so we do not treat them as beings with intrinsic moral value. I cannot relate to my Roomba - we cannot look one another in the eye, have a conversation, or express physical affection or rejection. There is movement toward understanding what it is in other beings that elicit these kinds of reactions from humans - cuteness, responsiveness, etc.[435] To the extent that these features have been built into robots, humans have tended to react to them in emotionally significant ways, even when their programming is otherwise fairly elementary.[436]

---

[435] Cynthia Breazeal, *Designing Sociable Robots,* (Cambridge, MA: MIT Press, 2002), 7-10.

[436] This is particularly true of robots like Boxie, which exhibits cuteness and a toddler-like helplessness in order to elicit human empathy. http://labcast.media.mit.edu/?p=206

In order for a robot to be fully understood and treated as a person with moral and theological value, that robot would probably need to have at least some sense in which it showed the humans with whom it had relationships that it had a sense of itself and its place in the world. Its actions would indicate a questioning of the world, a drive toward understanding moral meaning, an inclination toward goodness and a freedom to pursue that goodness. It, as a human-created product, would also exhibit the downsides of being derived from finite, situated humans: such a robot would also be able to understand what a lack of goodness, lack of meaning, and lack of freedom would be. Like chimps, they could lie and be unnecessarily aggressive, simply because they chose to do so. It may be that it is primarily in this latter category of actions, and not the former, that we truly come to recognize humanoid robots as being "like us."

Relation to Human Beings as Objects of Moral Action

Up to this point, I have been describing a potential theological understanding of varying layers and types of personhood - how "we" ought to treat "them" (and "us") and why different beings may be best treated somewhat differently, even within the schema of beings that have intrinsic value. But certainly not all of "us" actually act that way, or act with the same reasons. A descriptive account of how various human persons behave toward other human persons and other beings of intrinsic moral value must be more varied, as there are human beings who do not treat other humans humanely, much less chimps, cats, or robots. Part of the fallenness of humanity includes our ability and tendency toward inhumanity - through violence, murder, rape, and the like - but also

through ordinary acts of thoughtlessness and neglect. But there are also some tendencies of human behavior to treat that which appears to be living and human in certain ways as though it actually had these qualities.

Cynthia Breazeal, in her work in designing sociable robots, explains various aspects of sociable robots that lead most people to treat them as though they had minds and could interact in genuinely sociable ways. Her understanding of robot sociality arises from psychological studies based in evolved human biology, that "the human brain evolved in a world in which only humans exhibited rich social behaviors, and a world in which all perceived objects were real physical objects. Anything that seemed to be a real person or place was real."[437] That is, robots that appeared to be aware of the humans around them and responded to them in ways that indicated empathy, emotion, and social recognition were most likely to be treated by humans as though they were persons.[438] The more we recognize our own behaviors in robots, and more so in emotional and empathetic interactions than intellectual ones, the more we are likely to treat them as fellow persons of moral status, with needs, emotions, and interests of their own. In this sense, we are not only more likely to treat them as worthy objects of our own moral action, but also to ascribe to them moral subjectivity.

Robots as Subjects of Moral Action

In identifying some of the various qualities that both tend to lead humans to treat

[437] Breazeal, *Designing Sociable Robots,* 15, emphasis omitted.
[438] Ibid., 7-10.

robots as morally valuable objects and that give us reason to treat non-human animals as beings of (at least some) intrinsic moral worth, we may now begin to consider the characteristics that might give rise to an understanding of humanoid robots as the subjects of moral action toward themselves and others. In this section I want to argue that three attributes of personhood identified in chapter 3 in constructive futurist thought - embodiment, sociality, and situatedness - tend to mark those created beings whom we can and should identify as persons, but that there are other theological characteristics that provide a more complete picture of the fullness of personal life. All created beings have these qualities in varying degrees throughout life, and they are not absolute constraints, but ways in which we can reflect on which nonhuman beings might be given a "presumption of personhood" and why. Finally, I want to attempt to identify what may be said of nonhuman persons that could be considered a creation in the image of God, or if such a category is either unnecessary or unhelpful in marking these non-human persons as beings of intrinsic moral worth.

Three Futurist Concepts: Embodiment, Sociality, and Situatedness

I have argued for the idea that living creatures that are embodied, sociable, and situated in their environment have the basic qualities that allow us to consider them as beings of intrinsic moral worth and proper objects of human moral action. I believe that these are also necessary conditions of possibility for these beings to become subjects of moral action. These three categories form the ground upon which moral action may be built and from which other necessary conditions arise. That is, one must be embodied,

sociable, and situated in order to have the possibility of developing a moral sensibility that is lived out in the context of created life.

First, in order to become a subject of moral action in the physical universe, one must be physically embodied in some way to impact that universe. This may take a wide variety of forms and may include remote types of action, but the moral actions of care, compassion, and justice need the moral subject to be embodied in the world. This embodiment need not be carbon-based and can include different states of mind, intelligence forms, and learned behaviors, but it needs to exist in some way. As was indicated in chapters 3 and 4, embodiment is not simply an optional characteristic of personal life, but inherent to it.

Second, a subject of moral action must be sociable in order to behave morally toward others. No one learns or lives in a vacuum, and no one can behave ethically in a vacuum either. We learn with and from one another, and a subject of ethical action must have some kind of object for that ethical action. Just because a being is sociable does not automatically mean that it is an ethically-capable being, but a being must be sociable in order to be an ethical subject. This does not mean that one must be an extrovert in order to be ethical, but it does require that one live in relationship with others in order to pursue the good with and for oneself and the other.

Third, one must be a situated actor to be a moral actor. We each live in a particular time, place, history, culture, and tradition of learning, and it is from these particular places that our moral actions arise. We transmit what we learn, and our creativity moves in the situations and systems we inhabit. Just as there is no disembodied ethical action and there

is no relationship-free ethical action, there is also no ethical action or actor that is not finitely situated. There is no universal standpoint from which creatures can act, ethically or otherwise. A being that has embodied, particular relationships grounded in history and culture is in a position from which to act ethically.

Again, this does not automatically mean that all embodied, sociable, situated beings are ethical subjects. A cat has these characteristics and can be an object of moral action, but she has no sense of ethics from which to act ethically. She does not understand what is "right" and what is "wrong," only what leads to pleasure and pain and what builds or injures the relationships that she has with those around her. She knows that her humans are upset if she eats the food from her brother's bowl, but she does not have a sense that she is doing "wrong" to the other cat in taking his food. The chimp, on the other hand, has at least a rudimentary sense of what is owed to himself or another in a situation of limited resources. He knows that he can hide a piece of fruit from the others in order to avoid sharing but that if he is caught the others in his group will be angry that he lied to them. He has a sense of altruism toward others in the group, giving resources at some times in order to receive resources when he needs them. He experiences loss when another chimp dies and expresses something like mourning. In some ways, it is through the violation of ethical norms among chimps that we know that they have a sense of those norms. The ways in which chimps punish violators of an expected social code - those who hide food, who pretend to be nurturing only to bite, who steal from others - are subjected

to social, sexual, and nutritional sanctions from the larger group.[439] Chimps' social situatedness in a particular chimp "culture" offers them a set of norms to which they are expected to conform and learn from. But more is needed for a robust sense of what it means to be a moral agent or actor in the world.

An Irreducible Who-ness: Theological Features of Created Persons

It is at this point that I believe that three sets of theological qualities (which appear as paradoxes) are useful in developing that robust sense of what it means to be a moral actor and a full person in the created world. As mentioned above, it is not that beings who lack any (or all) of these qualities then also lack all intrinsic moral value; rather, it is that these qualities are necessary for one to be a complete ethical subject and to embody the completeness of what it is to be a person, morally and theologically. But in addition to the qualities of embodiment, sociality, and situatedness that are used in contemporary robotics as well as theology, one can add freedom and constraint, goodness and fallenness, and finitude and transcendence to the possible picture of personhood and moral subjectivity.

Freedom and Constraint

A being who experiences itself as having both freedom and constraint in its choices, actions, and relationships is one who has the potential to be a moral actor, as moral action requires some form of freedom to choose to be moral. Among creatures, this

[439] de Waal, *Good Natured*, 76-77.

also means that one's choices are not unlimited; they are constrained in many and varied ways. Because of our creaturely situatedness, we are born into systems and relationships not of our own choosing that constrain all of our choices in life. For instance, we cannot simply will a peaceable world into existence. Even if my "will" wanted that peaceable world, my powers and freedoms are significantly constrained in implementing that desire. But I can act ethically in such a way that may lead to that more peaceable world. I can choose actively, daily, to live peaceably with others; I can attempt to encourage peace among the communities of which I am a part; I can make peacebuilding an active aspect of my volunteer or professional work. I have some real freedom to choose my own actions; without this freedom, I cannot be truly responsible for my actions. But absolute freedom is neither necessary nor possible for creaturely ethical subjects. Having some real freedom is a necessary foundation on which to build that ethical subjectivity. In order for a humanoid robot to become an ethical subject, it would need to have freedom in a real sense, to choose to pursue the good in a given situation. Its choices will always necessarily be constrained, just as human choices and actions are constrained, but having an internal ethical calculus and drive toward the good, and the freedom to pursue that good, are critical factors in determining the ethical subjectivity of a machine as well as a living being.

It is again helpful to balance the theological views of Hefner and Waters, and to reframe the discussion through a feminist lens (here, through the work of Kathryn Tanner and Katie Cannon) in order to arrive at a more robust understanding of the rich paradox of human freedom and constraint. On the question of freedom, Hefner acknowledges that

human life is conditioned by our evolutionary heritage in many ways, but he still wants to argue that the essence of human life - all that one needs to know about the human condition - is freedom. Although he allows for conditionedness, Hefner does not take seriously the extent to which human life, choices, actions, and limits are constrained. On this, Waters provides an important counter-perspective, with his emphasis on the limitations of human life. But Waters does not take seriously the real forms of freedom that are both inherent and necessary in human life. His focus on the acceptance of limits should be challenged where it encourages human persons to be apathetic in accepting a status quo of injustice and freedom wrongly denied to persons.

We need not take a hard and fast position on the classic question of whether human beings have free will or whether God has preordained everything that happens in creation from before time. Christian theology has wrestled since at least Augustine about how human freedom and divine freedom of action relate to one another. Traditionally, the relationship between the two was posed as an unavoidable dilemma: if God truly ordains and orders the world, then God knows and determines all that will ever happen in history, then human freedom is void and God is responsible for our sin; but if human freedom is real and we are responsible for our own sin, then God cannot truly be sovereign, and then neither the promises nor grace of God can be assured. The former is a theistic version of the hard determinism posed by Minsky and Hall (and in a different way, Kurzweil) within the context of robotic futurism and has been offered by some theologians as the best way to read the universe. It makes God sovereign but not ultimately good. The latter offers a perfect goodness of God but makes God something less than God - the God who

promises the world and hopes for the best, not really being able to live up to the covenants that God makes and redemption that God promises. Neither offers a truly satisfactory solution, so it is necessary to reframe the problem.

Contemporary theologian Kathryn Tanner offers a different way out of the problem that involves a different understanding of the relationship between divine and human freedom so that the two do not dissolve into a zero-sum game, in which more of one means less of the other. Her development of the "non-competitive relation between creatures and God" begins with an understanding of non-competition between the divine and human natures in Jesus.[440] In this, it is "better to think of divinity and humanity not in terms of isolable, discrete quantities that divide up Jesus' life and persons, but as what characterize Jesus' life overall, or as a whole."[441] Tanner avoids a hard divine determinism, then, by explaining how "relations with God are utterly non-competitive because God, from beyond this place of created reality, brings about the *whole* plane of creaturely being and activity in its goodness."[442] Our real moral agency is not diminished or eliminated by God's working in the world, but is rather facilitated by it; it is one of the gifts given by God to us.[443]

Assuming, then, that both divine and human freedoms are real on some level, we may then consider the constraints on human freedom within the created world. Where classical theology spoke of the freedom of the "will," contemporary theologians,

---

[440] Kathryn Tanner, *Jesus, Humanity, and the Trinity: A Brief Systematic Theology*, (Minneapolis: Fortress Press, 2001), 2.
[441] Ibid., 16.
[442] Ibid., 4.
[443] Ibid., 70.

particularly those within liberationist traditions, have understood freedom to function on a variety of levels and a variety of ways. Katie Cannon particularly highlights the significant constraints under which African-American women have been forced (by their fellow humans) to operate, and how freedom works under those conditions of constraint. Both freedom and constraint operate on one's available choice of actions, as well as freedom and constraint to implement one's choice of actions. Where "dominant ethics assumes that a moral agent is to a considerable degree free and self-directing," womanist ethics serves as a corrective to recognize the limits and constraints placed on that freedom, and the spaces for freedom within deep constraint.[444] For Cannon, womanist "ethical analysis distinguishes between 'possibilities in principle' and 'possibilities in fact,'" refusing to ignore the very real structures of social oppression that put greater and different constraints on some people's actions and choices than others.[445] Under slavery, black women were subjected to all manner of constraint, dehumanization, and torture, restricting the freedom of their bodies for work, childbearing, learning, travel, food, and shelter. Within these massive constraints, very different from those borne by those who enslaved them, they carved out spaces and made their freedom. They were victimized, but they did not allow themselves to become only victims; they were constrained but were not defined only by these constraints.[446] Christian theology offers a picture of human life and moral agency that is defined in the tension between freedom and constraint in the

---

[444] Katie G. Cannon, *Black Womanist Ethics*, (Atlanta: American Academy of Religion, 1988), 2-3.
[445] Katie Geneva Cannon, *Katie's Canon: Womanism and the Soul of the Black Community*, (New York: Continuum Press, 1995), 125.
[446] Ibid., 49.

context of the goodness and fallenness that characterizes human moral systems of action.

Goodness and Fallenness

Likewise, ethical subjectivity among creatures also exhibits the characteristics of both goodness and fallenness. Within the created world, sin exists. It exists within every aspect of human life, and we are born into a situatedness that is characterized by fallenness. We are born into and live in a world that is not always just, not always loving, not always gracious, and so we ourselves grow and develop in ways that reflect the environments we live in. We, too, are fallen. But there is also real goodness in the created world, and ethical subjectivity requires a sense of that goodness and a desire for its pursuit amidst the fallenness of the world. We are all fallen, but we are never only fallen. God created the world as good and gave human persons a sense of and desire for that goodness. A robot who is an ethical subject must also understand the difference between good and evil and have a desire for the good (at least much of the time). Like most people, it may not have a highly developed system of ethical theory, but it must at least have some form of "commonsense morality" and an internal ethical differentiator. It need not, and indeed cannot, be an ethically perfect actor. No human product could be ethically perfect. But it should have some genuine sense of the good and be designed to strive for that good in a variety of changing circumstances.

Again here, Hefner and Waters each tend to emphasize one side of the paradox over the other, though both theologians certainly uphold that we created persons are both good and sinful. They also tend to emphasize particular forms of sinfulness and goodness

over others and end up with an unbalanced picture of human life. In Waters' work, we are created good by God, and our finitude is part of that. Our sinfulness arises primarily when we seek to overstep our God-given limits and reject our finitude. For Hefner, we are good creatures who are created by God in order to be creative, and we sin when we refuse the gift of creativity or use it to harm the world and its creatures instead of improving the world in which we live. For each, sin is primarily individual in nature but can have structural implications, though neither examines the structural aspects of sinfulness in any detail. Their views of both the goodness and fallenness of humanity are limited to the single aspects of human life they find most important, whether creativity or finitude. Here also we may say that each provides an important insight into human life, but they are stronger when both views are upheld together. We are created good by God and are given both our finitude and our creativity. We are called not only to accept our finitude and its attendant limits, but also to care and to alleviate suffering where we can. Because the world we live in is fallen, we cannot simply read the good off of nature, nor can we assume that the structures of human and created life are any more free from fallenness than we are. Our goodness and our fallenness are both multifaceted and take on different forms in different times, places, and people.

On the one hand, we were created by God and called good; the creation (including humanity) has not lost its goodness or redeemability. Humanity has a marvelous gift for goodness, truth, and beauty in the world that has not been lost. We have moments of redemption in which we reach out and make life a bit better for someone else. We create great art and literature and express the joy of life in God through worship that allows us to

reach beyond ourselves. And yet we also find ourselves continuously falling into our own sin, whether individual or systemic. We are caught up in systems of sinfulness, trapped into cycles and patterns of behavior that injure and afflict. We are greedy, prideful, self-destructive, and wasteful. We kill one another and destroy the very creation that we inhabit. Christian theology can both accept and offer an explanation for the tension in this facet of human life and ethics.

Our goodness, given by God and for which we are being transformed in the redemptive work of Christ, is created and designed for us. Whatever else we may be, we are the good creations of God. That is, it is good that we exist, as God desires us to exist and to be in relationship with us. This is quite different, though, from any sort of guarantee on our ethical choices. We have been created in goodness for goodness, and so the pursuit of that goodness is natural and inherent within us. This goodness within us participates in the inbreaking of the reign of God in the world; it is restored in the liberatory process of redemption. Our hope is in the full re-creation and resurrection of the goodness for which the world was made. We yearn and strive for justice and peace in the world; our desire to experience transcendence is, at its heart, a desire for the fullness of goodness and beauty that exists perfectly only in God. We work and desire and hope for the state of integrity and wholeness which we know that we have lost.

Our fallenness, in which we discover the depth of our own imperfections and tendencies to injure, should give us both theological and practical reason to pause when evaluating the optimistic predictions of futurism. Our goodness is not the whole of who we are, nor is our fallenness the whole of who we are, though it touches all of who we are.

We are, to paraphrase Luther, simultaneously good and fallen, fully good and fully sinful. In Calvinist language, we are totally depraved, which refers extensively to the effect of sinfulness and not intensively.[447] Serene Jones argues that "no part of our bodies, intellect, or soul is untouched by sin; there is no pure essence within our being that escapes the need for grace - sin goes 'all the way down,' and…no one eludes its taint."[448] Our sin, our rejection of God, our loss of original integrity - these affect every area of our lives. This "undoing forcibly violates the integrity of an entity by taking away all that defines it, offers it structure, and maintains its internal coherence."[449] Our lives are disintegrated by sin, which appears as individual, interpersonal, and systemic brokenness. We are born into systems of sin from which we cannot escape, and we violate our own integrity and the integrity of others throughout our lives. This sinfulness is by no means the final word - grace steps in - but a Christian understanding of the depth and breadth of human sinfulness should cause skepticism regarding all utopian projects or claims.

Finitude and Transcendence

From this sense of the good and a desire for that good, then, lies the potential for a transcendence of oneself and one's interests. Robots, like all creatures, are finite and

---

[447] John Calvin, *Institutes of the Christian Religion*, ed. John T. McNeill, (Louisville, KY: Westminster John Knox Press, 1960), 291. Calvin takes a stronger tone here in which total depravity can also be read intensively - that there is absolutely nothing good in us. Calvinist theologians today tend to read it extensively - that depravity extends to every area of our lives. As an example, see Serene Jones, *Feminist Theory and Christian Theology: Cartographies of Grace*, (Minneapolis: Fortress Press, 2000).
[448] Serene Jones, *Feminist Theory and Christian Theology: Cartographies of Grace*, (Minneapolis: Fortress Press, 2000), 102-103.
[449] Ibid., 103.

subject to vast limitations. One thing that currently makes humans unique is our sense of a world beyond ourselves and our present circumstances. We seek to overcome our own finitude, to yearn for more than our capabilities allow us, to have our reach exceed our grasp. This desire to transcend, to learn more, to grow and change is innate in the human spirit and part of our personhood. It is the part of us from which narratives of progress develop and from which we desire to change ourselves. It is the human capacity for transcendence that feeds the religious impulse and drives us to try to recreate ourselves through mechanical means. Our human desire for transcendence is the precursor to spirituality and the yearning for God, who transcends all creation. Our humanness allows us to desire to be more than we are, even as we recognize the limitations that are just as much a part of who we are. In a humanoid robot, this transcendence might first look like curiosity and a desire to explore the world beyond itself, not simply the larger environment in which the robot is situated, but the world beyond what it knows. The curiosity to understand and to explore could then develop into a desire to impact the world, to reach beyond itself into community and service, to reach beyond its limitations to make the world a better place. Finally, the robot might have the desire to reach beyond itself by changing itself, searching for meaning in its existence, and searching for a truth it cannot see with its actuated sensors.

The tension between these two poles of human life was identified by Reinhold Niebuhr as that which humans chafe against and come to fall into sin.[450] In our limitation

[450] Reinhold Niebuhr, *The Nature and Destiny of Man: A Christian Interpretation*, (Louisville, KY: Westminster John Knox Press, 1996), 182.

we experience eternity, yet in our yearning for that which is beyond ourselves we glimpse possibilities for more than we are from within what we are. We were made to long for more, to learn and grow, to desire transcendence; our quests for immortality and eternity are a part of us just much a part of us as our inability to reach them.

Where Waters emphasizes the "finitude" side of this paradox, Hefner implicitly emphasizes the transcendent aspect of human life. Waters' strong consideration of human finitude is necessary to keep our more grandiose conceptions of ourselves in check, but he ends up unintentionally making finitude and its attendant sufferings values in themselves that have no correlate transcendent mode, which leads him to reject human creativity and (to an extent) technology, including their capacities to alleviate the more problematic aspects of human finitude and suffering. Hefner, on the other hand, with his emphasis on human creativity, elevates the transcendent aspect of human life and downplays (though does not eliminate) human finitude. Because he considers most strongly our capacity to create and design, Hefner sometimes fails to take seriously our finitude and the very real limitations it imposes on us. Rather, to avoid the problems of these extremes, it is theologically and ethically important to uphold both sides of the paradox, both the finite and the transcendent aspects of created personal nature.

It is at this point that an embodied, sociable, situated robot becomes a spiritual being. As the robot searches for meaning it may find a sense of the divine and of meaning beyond the empirical world that it can see and sense. At this point, the personhood of the robot cannot and should not be denied, for it exhibits all of the personal qualities that mark human beings as "personal." The presumption of personhood may arrive before this

- perhaps when a robot develops a sense of the good and begins to question the justice of the present world - but at least now, the robot should be treated as a full person. Robotic spirituality may look very different than human spirituality, and it may be just as diverse in its forms as human spirituality. Robotic persons may be monotheists, polytheists, pantheists, or nontheists; they may have a mechanomorphic, rather than an anthropomorphic, deity.[451] Robotic spirituality and quest for meaning, transcendence, and the divine thus becomes the apotheosis of its personhood. This embodied spirituality arises from a robot's basic embodiment and social character, its freedom within constraint, its desire for goodness in a fallen and unjust world, and its sense of transcendence and meaning beyond itself. When a robot desires God, we cannot but admit that it is a person.

Image of God

Could a robot rightly be said, as humans are, to be made in the image of God? That human beings are "made in the image of God" is one of the most cited, and least understood, claims about the theological and ethical status of human beings. It is claimed in the creation narratives that humanity was created in the image of God, but this claim is rarely repeated anywhere else in scripture, nor is it given any specific theological content.[452] What this "image" consists of is left entirely open. So theologians and others have sought to fill in that content, with a wide range of results over the centuries. While a

[451] An archetype of a differentiated robotic spirituality that differs from its human counterparts is shown in the 2004 television version of *Battlestar Galactica*, in which the Cylons are primarily monotheists, though some are atheists, where the predominant human civilization is polytheistic.
[452] Genesis 1:27-28, Genesis 5:1-3, Genesis 9:6, I Corinthians 11:7, James 3:9.

comprehensive account of the choices that have been made regarding the content of the image of God cannot be made here, there are several broad strands that have been considered, including rationality, relationality, moral awareness or freedom, Christology, and image as empty category.

The first of these, rationality, is the classic answer arising within the Hellenistic context of early Christianity and remaining popular over the centuries. Its basic claim is that what is imaged of God in humanity is the rational mind or intelligence.[453] God is the most supremely rational being, so the argument goes, and humanity is distinguished from other creatures by our innate intelligence and rationality, so within this difference must lie the content of the "image of God." Because only human beings are said in Genesis to be made in God's image, and this is what makes us different from other creatures, this is what the image consists of. Rationality is generally what has been considered most valuable within human beings, as remains true within the dualistic tendencies of robotic futurism. We are given dominion over the rest of creation because we have the intelligence (and thus both the right and responsibility) to govern all other creatures. Within the "great chain of being" humans rank toward the top because we are among the most rational, underneath only God (who is perfectly intelligent) and the angels (who have perfect, though finite, intelligence). Our knowledge is not perfect, but the quantity and quality of the knowledge we possess puts us ahead of other creatures. It was also historically used to defend a hierarchy among human beings: educated white men being the most rational,

[453] The same argument was occasionally used in relation to the "will," though the arguments run roughly the same, with the same results.

and women, non-white men, and poor people farther down the line.[454] Aside from the historical and nearly inherent ethical problems of the rationality view, there is no scriptural evidence to support it as the content of the image of God. Reinhold Neibuhr argued that "it has been the mistake of many Christian rationalists to assume that this term is no more than a religious-pictorial expression of what philosophy intends when it defines man as a rational animal."[455] Simply put, nowhere is it said, or even implied, that rationality or intelligence is the basis of human beings' status in the image of God, nor is the intelligence or rationality of God spoken of in this context.

The second concept of the image of God, which is fairly prevalent today and which covers a variety of understandings of the content of the image, is that of relationality, or the ability to develop and sustain loving, caring relationships with other creatures and with God. Like rationality, this is another capacity-oriented concept of the image, and it falls prey to many of the same flaws. In this understanding, we image God when we nurture relationships, first and primarily with God and secondarily with other people. The "relational turn" in theological anthropology has made this a popular option in the past century or so, as its proponents seek to avoid the hierarchicalization of the image that plagued the rationalist view.[456] But just as with the rationalist view, there is not a strong basis in the reading of the narratives in which the idea of the image occurs to provide this kind of content.

---

[454] Cannon, *Katie's Canon*, 50.
[455] Niebuhr, *The Nature and Destiny of Man*, 13.
[456] F. LeRon Shults, *Reforming Theological Anthropology: After the Philosophical Turn to Relationality*, (Grand Rapids: Eerdmans Press, 2003).

A third option involves some other important form of capacitation, including that of moral awareness or freedom, is similar to the second in both its basis and ethical aims. In the case of moral awareness, to be a human being in the image of God is to have a sense of the good and of God. Here, what is unique about human beings is our God-consciousness and/or ability to choose between good and evil.[457] Because this is regarded as a particularly theological conception, and as it justifies an understanding of the image based not just on a particular chosen capacity but the way we exist in an aware and redemptive relationship to God, it comes a bit closer to the scriptural accounts, not in Genesis, in which the image is given no content, but in Colossians 1:15, in which Jesus Christ is declared to be the image of God. In Schleiermacherian terms, then, if Jesus was the image of God and had a perfect God-consciousness throughout his life, and we are made in the image of God, then we image God to the extent that we are molded to the image of Christ in reflecting this God-consciousness. We image God when we do that which is good, when we pray, and when we are filled with grace. In Hefner's conception, the significant capacitation is that of freedom, by which we create in an analogous way to the creativity of God. But this is again something of a stretch in claiming what is in the text. Colossians 1:15 does not say in what way human beings are "in the image of God," it simply says that Christ himself *is* the image of the invisible God.

Fourth, some modern theologians have claimed that the image is nothing more than God having chosen to be in relationship with human beings. There is no capacity in us

---

[457] Friedrich Scheiermacher, *The Christian Faith*, H.R. Mackintosh and J.S. Stewart, eds., (New York: T&T Clark, 1999), §91. J. Wentzel vanHuyssteen, *Alone in the World? Human Uniqueness in Science and Theology*, (Grand Rapids: Eerdmans Press, 2006).

that can be considered the image, only in God within us. This reflects in some ways the chosenness of Israel: Israel was not chosen by God to be a blessing to the nations because of anything particularly unique or special about Israel; rather Israel is special because it has been chosen by God.[458] So also, in this way, the "empty image" thinking goes, human beings are made in the image of God because God has chosen us for a redemptive relationship.[459] But this understanding goes both too far and not far enough in explaining the scriptural references to the image of God. On the one hand, it fails to explain why humans are in the image of God in a way that other creatures (and the creation itself) are not if indeed the whole of creation stands in a redemptive relationship with God. It is claimed that all of creation is groaning in waiting for redemption by God, and it is left unexplained here what it is about the divine-human relationship that makes it uniquely in possession of an "imaging" relationship, or why the concept of the image is used at all instead of something like chosenness.[460] On the other hand, the idea of the image being solely a matter of chosenness tends to reduce what is important in human life to the divine-human relationship. "Since the whole human being has been made for such a relationship, little interest need be expressed in particular characteristics or dimensions of their nature that distinguish them from other creatures."[461]

Finally, a Christological conception of human beings made in the image of God begins with Colossians 1:15 as the starting point from which to understand the image. In

---

[458] Deuteronomy 7:7-8.
[459] Karl Barth, *Church Dogmatics III/1*, ed. G.W. Bromiley and T.F. Torrance, (Edinburgh: T&T Clark, 1958), 183-187, 289-295.
[460] Romans 8:22.
[461] Kathryn Tanner, *Christ the Key*, (New York: Cambridge University Press, 2010), 2.

this concept, one begins by reflecting on the fact that Christ is himself the image of the invisible God, and we are made in that image, so we understand ourselves as humans properly by reflecting on Jesus as the perfect image. Reinhold Niebuhr argues that a key "characteristic of the Christian view of man is that he is understood primarily from the standpoint of God, rather than the uniqueness of his rational faculties or his relation to nature."[462] In this sense, Kathryn Tanner suggests, "a Christ-centered treatment of our creation in the image of God turns attention away from the human altogether."[463] At this point, the Christocentric understanding of the image of God can be taken in at least three very different directions. First, it can then be claimed that human beings are in the image of God to the extent that we image Christ in our lives. This results in a graded concept of the image, in which the holiest among us are truly "in the image of God" and those whose lives and relationships are less like those of Jesus have a substantially more marred image. It can even be claimed that those who have not experienced redemption in Christ (as defined by baptism or the like) are not in image at all, as they have not been transformed and remade into the image of Christ. The Christocentric definition of the image then becomes a means by which to divide the human family and to set up a moral or theological hierarchy of those who bear the image and those who do not. The fundamental moral and theological equality of all human beings before God can be easily lost here. In a second option, that taken by Waters, the image of God is Christocentric in something like a kenotic model, in which an acceptance of our weaknesses, frailties, and finitude are

[462] Niebuhr, *The Nature and Destiny of Man*, 13.
[463] Tanner, *Christ the Key*, 1.

modeled perfectly in Christ. It is thus when we accept our own finitude and mortality, embracing the necessity of life and hope in the redemption in Christ, that we become transformed into the image of Christ and disclose our imaging of God.[464] He also ties this with an emphasis on chosenness and the lack of any inherent capacity for humans to image God; in order to become the images that we are, we must be transformed in Christ. But the Christocentric understanding of the image can also be read in a very different direction that need not divide humanity once again into the elect and the reprobate. We can claim that Christ is the perfect image of God, divinity and humanity perfectly united, and the shape toward which human life should be made. Within this, though, all human beings are equally corrupt, equally in need of redemption, and equally chosen by God for that redemption. The image here is a reflection of a Barthian understanding of election - just as Christ is the elect one in whom all humanity is elected, Christ is also the image in whom all persons image God.[465] Christ is the image of God on our behalf and in whom we "become who we are" before God. We are transformed by Christ, and the redeemed life reflects the inbreaking of the reign of God, the already within the not yet. Tanner goes a bit farther to claim that "What is unusual about human beings - and what therefore makes them in the image of God as other creatures are not - is that the character or identity of human life is remolded in the process. Humans do not simply reflect the image of God. In so doing something happens to human life itself. Its very own character is altered or

[464] Waters, *This Mortal Flesh*, 183.
[465] Karl Barth, *Church Dogmatics II/2*, ed. G.W. Bromiley and T.F. Torrance, (Edinburgh: T&T Clark, 1957).

transformed for the better."[466] We are in the image of God precisely in our transformability toward the image of God. If this transformability to become more like Christ stands as the content of the image, then it is possible for a non-human creature that was likewise able to be transformed into an ever-greater likeness of Christ to also bear the image, to be a (creaturely) person. What this likeness of Christ might be and how non-humans could be said to be in the image of God will be explored in further in chapter 5, but for now it can be flagged that this theological transformability, universal among humans but seemingly unique to us (for now), is a possible option for understanding how a "what" might become a "whom."

Kathryn Tanner's understanding of the Christocentric image of God developed as human transformability into the image of Christ offers one way to think about what it might be for a robot to be thought of as being "in the image" of God as well.[467] Given the relative emptiness of the content of the language of the image of God in scripture, we must take clues as to what the image might refer to in how it is used. In the New Testament, it is Christ who is spoken of, not as being *made* in the image of God, but actually *being* the image of God incarnate in the world. From this we can learn that to be in the image of God is to be in the image of Christ, God made flesh, and where we come to image Christ we live truly into the image of God in which we are created.

Tanner then uses this idea of being molded in the image of Christ as a model for

[466] Tanner, *Christ the Key*, 16.
[467] Ibid.

understanding human nature.[468] Human uniqueness, for her, thus consists in our potential for transformability into the image of Christ. Human nature becomes a matter of irreducible openness and changeableness, rather than a specific ability like reason or relationality. When we allow ourselves to be transformed, morally, theologically, and behaviorally into the image of Christ, we become ever more in the image that we were created to be. In this sense, then, a fully humanoid robot that had its own spiritual life and longing for God could potentially also image God in allowing itself to be transformed into the image of Christ. This possibility is far off - no currently existing or attempted robot has its own sense of God - but it is not theoretically impossible. A robot might even have a different sense of God, or what it means to image Christ, than human beings do, and it may image Christ differently, in the situatedness of its own existence. It could image Christ's incarnate life in what it chooses to learn, how it interacts with others, and how it uses its own power to transform the world around it toward God's redemptive purposes. The nature and shape of this robotic image is undetermined, but consideration of the possibility of a robot coming to be in the image of God as it allows itself to be transformed into the image of Christ could open possibilities for both robotics development and interactions with human beings as we reach out to the robot companions around us.

Impact and Future Possibilities for Human and Robotic Personhood

Theology and the Politics of Technology

[468] Ibid., chapter 1.

The Christian community has struggled with its relationship to the political systems in which it has found itself over the course of the centuries. In this set of theological reflections, I hope to elucidate some of the broad features of Christian political reflection that may be brought to bear in considering the systems in which humans plan for, design, develop, and utilize our machines, including humanoid robots, how we might widen the circle of grace for the flourishing of all. This is by no means a comprehensive account of Christian politics - such a thing would be impossible. Nor is it a claim that there is only one "real" version of Christian political understanding. While I do believe that there is room for substantial disagreement within Christian communities over the shape of political life, and that the relation of the Christian community to the political system as a whole varies greatly with time and place, I also believe that there can be such a thing as a real Christian social ethic that is not solely a reflection of the mores of the day.[469] In this, I seek to highlight some of the key insights and tensions that are of particular interest in the current context of technological development.

Skepticism About Privilege and Power Relations

In contrast to some futurists' claims of technoutopianism, the Christian tradition claims a skepticism about the ability of all human endeavors to be free of greed, exploitation, and selfishness. Christianity recognizes that we are not only good; we are also fallen, and this is reflected in all of our personal, interpersonal, social, and systemic

[469] My position here is intended to stand in contrast to the claims of Ernst Troeltsch in *The Social Teaching of the Christian Churches*, (Louisville, KY: Westminster John Knox Press, 1992).

choices and actions. We are rightly skeptical when those who already have power propose ideas that they claim will solve all of the world's problems, and more so when they are willing to abide the suffering of others as the price of "progress."

Some robotic futurists have shown a rightful skepticism regarding the role of government coercion in the adoption or non-adoption of emerging technologies. No one should be legally required to incorporate particular technologies into her body, and with some basic protections for public and personal safety, should usually not be restricted from doing so either. But these same futurists tend to discount other forms of extralegal coercion that can have just the same impact through economic or cultural means. For instance, there is broad agreement that one should not be required by law to receive a direct brain-computer interface if one does not desire it. But let's say I want a job as a network engineer – should the company hiring me be allowed to require that interface as a condition of employment? Should I be required to have certain "upgrades" in order to attend a particular school, or to receive health insurance? Coercion must be taken seriously as more than a simple matter of governmental intrusion or non-intrusion, and simply saying that people can opt-out is insufficient. People need health insurance, education, and employment, and as we design the humanoids of the future, the rights of humans 1.0 need to be protected in a broader scope.

Greater intelligence is not the same as greater goodness, and we have little reason to believe that our inventions, however magnificent, will be any better than we are. The Christian community has the opportunity to highlight the vast power differentials that exist between societies and the moral costs of those differentials. The people who build

the microchips and other parts that advanced information technologies depend upon are often not the same people who have the financial resources to purchase and use those technologies. Likewise, the ethical calculus of war changes dramatically when one side can place robots on the field of battle instead of human soldiers and the other side suffers most or all of the human casualties. Promises by futurists of "trickle down" technology, longevity, and wealth ought not be believed when empirical trends move in the other direction. Christians should be skeptical of the absolute urgency claimed for the sexy high tech and the utter demonization of anyone who opposes particular technological developments or finds that perhaps more funding should be given to alleviate the crushing problems of a lack of clean water, housing, food, and healthcare currently faced by many in the world. Surely we can both provide enough food for all while developing nifty humanoid robots and the like. But we humans are not very good at feeling the urgency for too many policy priorities at once – some things are emphasized while others are de-emphasized, and I want to ensure that in the rush to develop the new and the cool that the problems of today are not forgotten. There is such a thing as moral progress within Christianity, but it is not assured within the trajectory of history. We find moral progress as we become more like Christ, though on earth we can never reach this goal entirely.

Concern for the Least: Seeing What (and Who) is Obscured

Christianity is rightly skeptical of privilege and is called to keep its eyes on those who are considered the least in the eyes of the world-at-large. We are called to care for the sick, the hungry, the prisoner, the widow, and the orphan. The early Christian community

placed a strong value on caring for those who were left behind by Roman society, even as an illegal, underground religious movement. So today we ought to consider how the policies and practices that are taken for granted in Western society affect the poor and those who lack ability to have their voices heard in the public square.

Many advancements in medical technology, including sterile surgery, antibiotics, vaccines, birth control, and the like, contributed to a dramatic increase in life expectancy in developed nations during the 20th century. Scientific and technological advancements were critical to increases in health and longevity on a global scale, but technology alone does not solve global crises. As with all new technologies, emerging robotics technologies are first available only to the very wealthy and then eventually are common enough to be more readily affordable for people in developed nations. But many of these technologies do not "trickle down" to the point of being affordable to the world's poorest, or even the less well-off in the US. It takes significant and intentional choices in political, economic, and social policies in order to make emerging technologies broadly accessible.

As an example outside of robotics, AZT and other HIV medications were initially terribly expensive and unavailable to most people suffering from AIDS. After they had been introduced and political pressure in the US strongly encouraged the company making the drug to lower its prices, AZT began to be covered by more Americans' medical insurance. But AZT only became available to the world's poor because of massive political pressure on GlaxoSmithKline to lower its prices prior to the drug's patent expiring in 2005. Here the technology was a necessary, but not sufficient, condition for the possibility of solving the global AIDS crisis. It also takes political will, public

pressure, an effective global distribution system, and a strong healthcare infrastructure to provide the care needed by so many.

Effective and inexpensive treatments for a wide range of other conditions, such as malaria, remain underfunded in both development and distribution, so there remain approximately 225 million cases of malaria annually worldwide.[470] Public pressure here has resulted in a substantial increase in funding for insecticide-treated mosquito nets and antimalarial medications, but these are available at levels far below what is needed in order to fully and effectively combat malaria.

In the example of malaria, technologies such as insect nets and antimalarial medications have been developed and are part of the solution, but there also must be public pressure for the political decision to make the eradication of malaria a priority on a global basis. Once this policy goal has been established over time and pubic leaders are held accountable for upholding it and funding it as a priority, then there must be the necessary technological development to create the conditions of possibility for an effective solution. Substantial public and private money must be spent wisely to develop the technology, but even the development of the right technology is not the end. There also have to be just economic and political structures through which to distribute the technology, such as a new generation of antimalarial medications. Malaria is something of a different case here because there is not substantial public need for antimalarial medications in the developed world to compete with the needs of the developing world.

---

[470] WHO World Malaria Report 2010, http://www.who.int/malaria/world_malaria_report_2010/malaria2010_summary_keypoints_en.pdf.

There remains, however, the need for access to effective treatments by a broad range of persons throughout the world as well as the infrastructure to produce and deliver those medicines. Each step, from the decision to make a particular challenge a political priority, to the choice to fund the research and development needed to develop the technology necessary to create the possibility of a solution, to the political effort needed to ensure that the developed technologies do not simply benefit the already well-off, indicates that the problems are rarely primarily technological in nature. Technology is part of the solution, but if delivered unjustly, can also become part of the problem by exacerbating existing inequalities.

Moving back into the field of robotics, we can imagine that a fully capable humanoid robot personal assistant is developed in the next decade or two. This robot could walk, talk, keep your schedule, send messages, clean your house, drive your car, and cook your meals. Such a robot would be hugely expensive to develop, so the first question is whether this type of robot should be sought through publicly funded research. Once a choice has been made to actively fund the development of this robot, it would be very costly to conduct the necessary research and development to have the robot be safe and effective to use. So here the initial political decision has been made to conduct funded research – to make robotic personal assistants a public priority, which generally indicates the choice to not fund some other area of research. Certainly a policy decision could be made to allot a much greater sum to scientific research in general, whether medical, military, or otherwise, but this would also require much greater infrastructure in education, facilities, and equipment for years before becoming a viable option. Given

current infrastructure and funding levels, the decision to develop a humanoid robot would require the defunding of some other policy priority – and public pressure would help to determine what that would be.

Once the decision has been made and funding provided over the extended period of time necessary for development, only then do questions of distribution and access arise. Our robot would have the potential to be useful to anyone in the world and would have universal demand accordingly. Certainly the robot would initially be available only to those who could afford its marked up retail cost – it would not be available to an average citizen. It would take some time for this robot to be accessible to the middle class of Western nations, and given the universal market demand, would likely not ever be made readily accessible to the majority of the world's population. Unlike AZT or antimalarial medications, the target market for this robot would not be people in the developing world, so it would be surprising if it were ever made affordable to them. It would remain the province of the wealthy and would have the potential to increase existing disparities. It would also be likely to exacerbate other problems by taking away many low-level jobs that humans now have, such as taxi drivers, cleaning service workers, office assistants, etc. The wealth needed to take advantage of this technology would become even more unavailable to these service industry workers as they lose their jobs. I believe that it is the job of the Christian churches to pay special attention to these kinds of changes and their effects and to advocate on behalf of those who are left behind, unemployed and underskilled, as current technology trends move forward.

Beyond Utopia and Dystopia

That we are both good and fallen, free and constrained, gives us a perspective beyond simple utopia or dystopia – things are neither as good nor as bad as they could possibly be, because within each human person dwells the capacity and tendency for both good and evil. Robert Geraci has documented the strong apocalyptic tendencies among roboticists, and whether they predict a utopia or dystopia, they believe that the end of human history as we understand it is coming very, very soon.[471] Christian theology has often addressed issues of apocalypticism and has found ways to temper the more problematic elements of it - ensuring that Christians continue to seek justice and peace in this world even while hoping for the reign of Christ at the end of history. Like contemporary technological futurism, Christianity has its apocalyptic elements and must look into the future responsibly within its community. There are those within the Christian family who try to "read the signs" and imagine that they know where the world is going, whether toward a vision of paradise on earth or a fiery Armageddon. These visions have often captured the popular imagination in times of crisis and change.[472]

This is not the end of the story of Christian involvement in the future of history nor the history of the future. Rather than imaging an end of the world of heaven or hell on earth, Christians can offer a way to live in the world as we know it today, even as it changes dramatically with the emergence of various technologies. We can balance our

[471] Robert Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*, (New York: Oxford University Press, 2010).
[472] Tim LaHaye and Jerry B. Jenkins, *Left Behind: A Novel of Earth's Last Days*, (Carol Stream, IL: Tyndale House Publishers, 1995).

eschatology with a "counter-apocalypticism" that directs our energies toward the pressing needs of the present while remaining aware of the trends moving into the future.[473]

Technology is Neither the (Ultimate) Problem Nor the (Ultimate) Solution

It is in this perspective that we can understand that our technologies are, as an extension of ourselves, neither the ultimate problem nor the ultimate solution; they are simply tools that we can use for both good and ill. New technologies often solve one problem only to create others; these problems then require yet other technologies and the choices made in developing and utilizing them.[474] A Christian technorealism about the uses and abuses of technology can be instructive. Technology is not precisely neutral on its own; as the product of human choices it has particular ethical and political valences. But most important technologies have many possible uses, often ones entirely unimagined by those who created them. Religious ethical creativity here can begin to see how technologies may be used for good or ill, how to steer the uses of technology toward the flourishing of all, and how to alleviate or avoid the abuses preemptively, whether in development, distribution, or use.

Toward the Ethical Development of Humanoid Robots

Given this understanding of the characteristics and qualities that contribute to a

---

[473] Catherine Keller, *Apocalypse Now and Then: A Feminist Guide to the End of the World*, (Minneapolis: Fortress Press, 2004), xi, 273-310.
[474] Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Journal of Evolution and Technology* 9, no. 2 (2002).

fully robust robotic personhood and the qualities of intrinsic moral value that arise along the way, we can now explore what it might look like to develop such robots in an ethical way that facilitates this development toward personhood. Here I want to take a practical, rather than a simply theoretical, look at the implications of this understanding in the three primary contexts outlined in chapter 1: labor replacement and human services, military and defense applications, and sex and companionship. This will necessarily be a cursory exploration; each will look like something different in its actual historical development, but I hope to provide something of a broad overview into some possibilities and potential difficulties within each area.

Labor Replacement and Human Services

Within the context of labor replacement, we can think about four different areas of labor that correspond to four different sets of human needs and the robots required (and desired) to replace them. First, there are particularly dangerous, dirty, or undesirable jobs, such as mining, nuclear cleanup, long range space exploration, landmine deactivation, or bomb removal. These range from fairly dangerous, with a high risk of death, to jobs that cannot be performed by humans. These jobs are good candidates for replacement (or at least augmentation) by robots, for the good of the humans involved in these industries. Some of them will be able to save human lives from dangerous tasks, while others will be made possible only with the use of robots, as humans cannot survive the tasks. But we would not want to then replace the humans with fully-developed humanoid robots who could be considered intrinsic moral objects in their own right. We would actually want

robots who were not aware of themselves and only performed the tasks that they were designed to perform. These comparatively "dumb" robots need not be unintelligent; indeed, they must be able to safely and effectively perform their tasks. But this capability need not and should not be paired with the kind of moral and spiritual awareness that would make the robots "personal."

The second group of labor replacement tasks include tasks that are not inherently dangerous for humans to perform, merely extremely difficult or impossible because of the particular manipulations, skills, or precision required. These would also be good candidates for robotic replacement of the human labor and include precision surgery and nanoscale manufacturing. In the case of the latter, the manufacturing involved is not possible using traditional human labor in order to implement the nanoscale technologies; it still requires human design and development, but the actual manufacturing itself would be performed by the robots. In the case of the precision surgery, robotic intervention to replace or augment the human surgeon's skills is needed for the surgical intervention to be successful. Robots here can save human lives, and the humanoid-ness of the robots can vary substantially, depending on the circumstances. Because they are not being used to perform dangerous tasks, some of these robots may be fully humanoid, but given that they are designed to be used solely for these very specific tasks, we may not want them to be, as this limits their freedom to choose their "careers" in ways that persons would not want limitations imposed upon them.

A third category of human labor that could, and currently is, being replaced by robots is in everyday human tasks, and the desirability of robots for this replacement

varies with the tasks, but they generally involve taking a reasonably desirable (if sometimes dull) job away from a human being and giving it to a robot. Such jobs include almost all kinds of manufacturing and service industry jobs like housecleaning, medical assistants, and taxi drivers. On the one hand, many of these jobs can be fairly easily replaced by robots (to the extent that they have not been already), and can free people from boring jobs. On the other hand, though, these take away important low- and medium-skill jobs from humans who need them and who have fewer options left for other employment. This is not a recipe for trickledown wealth for all persons, but instead may lead to mass unemployment and a rejection of robots in general. These robots are likely to save a great deal of money in the long run for the companies that implement them, but the human costs may be staggering, so there needs to be some alternative for people whose jobs are displaced by them. In order for the robots to perform these jobs long-term without complaint or rebellion, most of these manufacturing and service-sector robots should not be fully humanoid and have capabilities limited to their tasks.

A final category in which robots could be used to replace human workers is in creative professions. At this point in time, this is the least likely sector for robots to be used, as they cannot currently perform the tasks as well as a human can, or the humanity behind the tasks is irreplaceable. These tasks include musical, artistic, and creative writing tasks, as well as constructive research and scientific scholarship. A robot may be able to process information well, but robots cannot invent the entirely new, decide what new avenues of research to pursue, or create human art. A robot that was fully humanoid, however, in the sense of understanding transcendence and spirituality as well as it own

social and cultural location, might be able to do these things. In this case, it would not, strictly speaking, be a replacement for human creativity, but an addition to it - robotic art and music and scholarship could augment the human varieties in ways that contributed to both human and robotic culture.

Military and Defense

In the arena of military and defense applications for humanoid robots, much current research and development is emphasizing both the replacement of warfighters in the field of combat as well as military intelligence augmentation through ever-more-automated drones. Current military robots augment the human warfighting force and provide services that humans cannot safely perform, like landmine and roadside bomb detection and deactivation. There are several current and likely future applications for military robots that might each require different sort of robotic personhood for their ethical implementation. Nearly all military applications for robots are at least somewhat dangerous - the entire premise behind military robotics development is to save human (soldiers') lives and to allow tactical applications that are too dangerous or problematic for humans to accomplish.

In the area of unusually dangerous military tasks, the same concepts apply that are relevant for robots in dangerous labor replacement. They should be able to perform their tasks effectively without being too self-aware. Landmine deactivating robots, for instance, do not need to think about goodness or transcendence; they just need to deactivate landmines so that human beings do not die. The same applies to robots that have a high

likelihood of being captured; in some ways, the less they "know," the better. Robots (at least at this point) cannot be tortured for information once captured, though their data stores can be decrypted and analyzed.

Robots used for human warfighter augmentation, like the Packbot or TALON robots, should also be able to perform their tasks well, though they may need a few other capabilities in order to effectively serve their human counterparts.[475] They should be able to call for help if captured or the humans with them are injured, and in more advanced applications, they may even be able to provide basic field medical care to assist injured humans. But as field augmentation devices, they need not be "personal" members of the squadron, and they probably should not be. If robotic augmentation devices themselves called for help when injured or the soldiers became emotionally invested in them, their presence could actually hamper the soldiers' ability to fulfill their mission in the field.

Intelligence-gathering robots, including more advanced versions of current drones, should also be able to perform their tasks without personal awareness or reflection. These robots should, tactically speaking, be able to strip themselves of the information and processing power that they possess if captured. A form of informational auto-destruct might even be helpful, and it would be important in those circumstances to not have a "personal" being captured or destroyed. Drones do not process the information they gather, and they do not make the subsequent tactical decisions based on that information, so as much as possible, they should be passive information gatherers that follow orders

[475] P.W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, (New York: Penguin Books, 2009), 26.

without reflection.

Tactical autonomous combatants, or soldier replacements, however, may need to take a very different path. To the extent that robots are ever given autonomous decision-making power over when and whether to injure or kill a human being, they need to be able to "think through" that decision with all the available evidence before taking lethal action. They must have sophisticated processors to distinguish combatants from non-combatants and to consistently follow the rules of engagement and the laws of war, without raping, pillaging, seeking revenge, or committing other atrocities against the civilian (or even combatant) population.[476] But the best ethical development of such robots would include not only the ability to distinguish combatants from non-combatants and weapons from toys. They should also be designed to defuse a chaotic and potentially violent situation so that killing can be prevented. They should be designed to use non-lethal means of halting combatant activity and to lower the civilian casualties of war to the greatest extent possible. While robots cannot ever "put a human face" on war, they can be designed to slow down and reverse trends toward violent escalation even within the context of war. They could, for instance, provide nutritional or medical assistance to non-combatants, even as they defended themselves from combatants. Human ethical and engineering creativity here could positively impact the war scenario and facilitate a quicker resolution to conflicts with fewer casualties on either side.

Unfortunately, this does not seem to be where trends in robotics development are

[476] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots*, (Boca Raton, FL: CRC Press, 2009), 29-36.

going, and it is important to note the changes to the entire conduct of war that arise when only one side has even the possibility of human casualties. Instead of finding ways to limit human casualties on both sides, the move toward robotic warfare is likely to increase both the likelihood of going to war and the ongoing pursuit of that war amidst high collateral damage if one side is fighting with robots and the other with human beings. Having robots in the field increases public acceptance of war, as one's children, friends, and neighbors do not have their lives at stake. Not having humans prosecuting the war in the field creates a scenario in which no one (on one side) sees the human face of the casualties on the other side. There is then little to stop the war from continuing longer than it needs to when neither side can "see" the other. Finally, robotic soldiers with autonomous fighting capacities are far more likely to be targets for interception and counterattacks in the field. They can (and likely would) be captured and reprogrammed to turn on those who sent them and could easily escalate the war in progress, endangering far more lives than before.

Love and Sex

As with robots for labor and military applications, robots designed for sex and companionship serve a variety of purposes and take a variety of forms. They have a unique element, though, in that the possibility for secondary effects in humans' behavior toward other humans is more likely to be altered because of an encounter with these robots. For instance, if someone had a humanoid companion robot, the person's interactions with that robot, good or bad, are more likely to affect the form and style of

that person's interactions with other human persons, based on expectations, emotional attachments, etc.

In cases in which humanoid robots are being designed and developed specifically for sexual applications, the line can become blurred between robot and human sex partners. Robots currently being designed for sex are intentionally built to have idealized feminine bodies, with large breasts, large lips, and spread legs.[477] They mimic an idealized female anatomy but do not have any other aspects of human women. They are designed, for instance, not to talk, make decisions, or have interests and ideas of their own. They are, in a very real sense, built to be sex slaves for the men who purchase them.[478] The men who desire such robots look for them, in many cases, as their primary sex partners, rather than augmentations to human partners, and they do so because they often have difficulty relating to human women.[479] Some argue that, for men who have difficulty relating to human women or who tend to be abusive to human women, a sex robot could be a good alternative so that actual human women are not harmed or otherwise negatively affected by these men. But I believe that this impact is likely to be the opposite, in desensitizing an even broader group of men to the needs and interests of human women because these robots are becoming available. Just as exposure to simulated violence in video games tends to make humans less sensitive to the harms of actual concrete violence, so also exposure to simulated violence or degradation in sex is likely to desensitize humans to actual harms

[477] *Sex Robot*, Discovery Fit and Health, 2011.
[478] This is a case in which the gender dynamics and specificity are important. There are no male-bodied robots designed for sex that I could find, and the client base for the female-bodied robots is more or less exclusively male.
[479] *Sex Robot*, Discovery Fit and Health, 2011.

to women in non-simulated sexual encounters.[480] As they interact with their robotic sex slaves, these men may be more likely to transfer their assumptions and patterns of interaction to real women, assuming that they are not there to talk or think, but only to provide sexual services. It more deeply ingrains the most misogynistic tendencies and further crystalizes gender assumptions, idealizations, and patterns.

Robots designed to provide both sex and companionship services are both somewhat similar to and somewhat different from purely sexual robots. Gynoid robotic "girlfriends" may have many of the same potential negatives and could either magnify these problems or help alleviate them, depending on how they are designed and used. Robots designed for both sex and companionship are likely to be subject to the same physical idealizations as those robots designed solely for sex, but they may also walk, talk, and interact with their human companions in more realistic ways. It is this realism that can have either a positive or negative impact on the individual owner and that owner's assumptions about and interactions with other human persons. As an example, if the companion robot were to be designed to interact with the person but not designed to disagree with him, have separate interests or ideas, or provide any critical feedback or expectations of its own, this could also instill in the owner the idea that human women are supposed to be "perfect" companions who simply do whatever men want them to do. As they interact with them as more realistic companions and have conversations and more human-like interactions with these robots, the danger of building false and damaging

[480] N.L. Carnagey, et al., "The effect of video game violence on physiological desensitization to real-life violence," *Journal of Experimental Social Psychology* 43, no. 3 (May 2007): 489-496.

expectations is even greater than with less interactive robots, in that a more realistic interaction tricks the brain into thinking that the interaction is real and can be translated into other patterns of behavior. On the other hand, in the less likely scenario that companion robots are designed and built to have interests and ideas of their own, to disagree with their human companions, and to provide the most realistic forms of humanoid interaction, they could help to disrupt these patterns and expectations among men who have difficulty forming relationships with human women. Perhaps the ultimate test would be whether a gynoid robot had the option to leave her human companion if she was being mistreated; in that case, the robot would not be a slave at all but would build expectations of fair and decent treatment into the humans that interacted with her. Such a robot would need to have a personal quality of understanding good and evil and desiring the good, not only in her own behavior, but in her own interests.

Other companionship-oriented robots are likely to have some humanoid qualities but not be designed for any sexual applications. These robots may provide care and companionship for the lonely, particularly children and the elderly. These robots are currently being designed and implemented, some in healthcare applications, others in classroom environments, and still others for home use as babysitters or pets. These robots could provide a much-needed service to those who lack sufficient human companionship, giving them an emotional outlet for caring and providing critical services when needed. A robotic dog given to an elderly person could not only be a caring (if self-sufficient) pet, it could also call for help if the human became ill. A child who is autistic could learn to interact with humans based on less emotionally-charged interactions with a

robot.[481] But there remain some dangers in the utilization of these robotic companions. They could be deployed to help lonely elderly people instead of providing needed human companionship, assuming that providing the robot is enough. Children with autism who learn to interact with robots might not then be given the subsequently needed therapy to help them interact with human beings. Where a robot is used to replace, rather than augment, human companionship, there is a risk of neglect and abandonment of the human who is in need of care. Even with a fairly robustly personal robot, this could be a problem if neither the robot nor the human are given the choice or chance to interact with others. Sherry Turkle warns of a second danger as well, that "robots that want to be buddies (are) implicitly promising an emotional connection that they can never deliver."[482] Robots that appear to be humanoid but are not seem provide an emotional bond with humans but do not actually do so, and these robots end up duping humans into emotional investments that are not returned and may take away from the emotional bonds with other humans. To the extent, then, that robots are not robustly personal, with interests, needs, ideas, and concerns of their own, they fail to live up to the promises that they offer as companions. If they were, however, to be developed in a robustly personal way, such that they could choose with whom to interact and how, having their own values and ideas, they could provide a type of companionship that is not based on an emotional falsehood.

---

[481] Chris Woolston, "Robots built to help autistic children," *Los Angeles Times*, October 17, 2011, http://articles.latimes.com/2011/oct/17/health/la-he-autism-robots-20111017.
[482] Jeffrey R. Young, "Programmed for Love: In a skeptical turn, the MIT ethnographer Sherry Turkle warns of the dangers of social technology," *Chronicle of Higher Education*, January 14, 2011, http://chronicle.com/article/Programmed-for-Love-The/125922.

Aristotle wrote that the highest form of human connection is that of friendship, so we can ask if it is possible that a robot could ever be a "friend" of this highest sort.[483] Regarding true friendship, he claims:

> That then is perfect Friendship which subsists between those who are good and whose similarity consists in their goodness: for these men wish one another's good in similar ways; in so far as they are good (and good they are in themselves); and those are specially friends who wish good to their friends for their sakes, because they feel this towards them on their own account and not as a mere matter of result; so the Friendship between these men continues to subsist so long as they are good; and goodness, we know, has in it a principle of permanence.[484]

Could this permanent sharing of goodness in friendship exist between humans and robots? As an ideal, it exists rarely among human beings in our own relationships. One can, Aristotle claims, only truly have one friend in this sense.[485] In order for such a relationship to exist between human and robot, the robot would have to be fully humanoid in the most robust sense - being spiritual, having a sense of transcendence, desiring the good both for itself and the other, and the human would have to be likewise. I believe that to the extent that such an Aristotelian friendship is available between humans, it is at least theoretically available between a human and a robustly humanoid robot. True friendship is rare, and true human-robot friendship would be rarer still, but there is not a technical reason why it could not exist, were a robot to become a person in the strong theological sense.

---

[483] Aristotle, *Nichomachean Ethics*, trans. D.P. Chase, (Mineola, NY: Dover Publications, 1998), Book VII.
[484] Ibid., 141-142.
[485] Ibid., Book VII.

Conclusion

A world in which humans and robots coexist is one with tremendous possibilities for good and ill. Futurist thought in robotics has contributed both positively and negatively to the development of humanoid robots to this point, offering ideas and values about what it means to be human and what it could mean for a robot to be a moral person. Some of the more popular forms of robotic futurism have tended to overemphasize intellection and a disembodied mind as the ultimate form of existence, while the more constructive forms have looked at human emotional and social interactions and patterned humanois robots after them. Robots that are embodied, sociable, and situated in their environment and history are ones that begin to mirror humanity and the beings that we consider to be morally valuable in themselves. But robotics and related psychology do not offer a complete picture into the possibilities for robotic personhood in interaction with human beings. It is here that theology can provide a useful history of reflection and understanding of personhood beyond the human that can begin to develop creative possibilities for the future direction of robotic personhood as well. Fully humanoid robots, then, could embody the qualities of freedom and constraint, goodness and fallenness, finitude and transcendence, and embodied spirituality that characterize human personal life. These qualities can be considered in the development of robustly humanoid robots in a number of different application areas and the ethical effects of those developments can be better understood using these criteria. Humanoid robots can perform jobs that humans cannot or would not do, they can change the ethical calculus of war, and they may even be able to provide genuine companionship and friendship to human beings,

but they should be designed in such a way as to facilitate human flourishing first, so that robotic flourishing can follow.

Bibliography

Ackerman, Evan. "Toyota's Healthcare Robots Are Ready to Help You With Absolutely Everything," *IEEE Spectrum*, November 4, 2011.

Agar, Nicholas. "Thoughts about our species' future: themes from *Humanity's End: Why We Should Reject Radical Enhancement*," *Journal of Evolution and Technology* 21, no. 2 (November 2010): 23-31.

Alexander, Brian. *Rapture: A Raucous Tour of Cloning, Transhumanism, and the New Era of Immortality*. New York: Basic Books, 2003.

Aristotle. *Nichomachean Ethics*. Translated by D. P. Chase. Mineola, NY: Dover Publications, 1998.

Arkin, Ronald. *Governing Lethal Behavior in Autonomous Robots*. Boca Raton: CRC Press, 2009.

Arnst, Catherine. "I, Robot: In developing robots to work alongside humans, scientists find even crude facsimiles of human behavior help people accept mechanical colleagues," *Bloomberg Business Week*, August 20, 2007. http://www.businessweek.com/technology/content/aug2007/tc20070818_479361.htm.

Association for the Advancement of Artificial Intelligence Presidential Panel on Long-Term Futures. "Interim Report from the Panel Chairs," Association for the Advancement of Artificial Intelligence, August 2009. http://research.microsoft.com/en-us/um/people/horvitz/note_from_AAAI_panel_chairs.pdf.

Asimov, Isaac. *I, Robot*. New York: Random House, 1991.

Augustine. *On Genesis*. Translated by Edmund Hill, O.P. and edited by John E. Rotelle, O.S.A. Hyde Park, NY: New City Press, 2002.

Bainbridge, William Sims. "The Transhuman Heresy," *Journal of Evolution and Technology* 14, no. 2 (August 2005): 91-100.

Barbour, Ian. "Neuroscience, Artificial Intelligence, and Human Nature: Theological and Philosophical Reflections," *Zygon* 34, no. 3 (September 1999): 361-398.

Barth, Karl. *Church Dogmatics III/1*. Edited by G. W. Bromiley and T. F. Torrance. Edinburgh: T&T Clark, 1958.

Bjork, Russell C. "Artificial Intelligence and the Soul," *Perspectives on Science and*

Christian Faith 60, no. 2 (June 2008): 95-102.

Bordo, Susan. *The Flight to Objectivity: Esssays on Cartesianism and Culture*. Albany, NY: State University of New York Press, 1987.

Bordo, Susan. *Unbearable Weight: Feminism, Western Culture, and the Body*, 10th Anniversary Edition. Berkeley, CA: University of California Press, 2004.

Bostrom, Nick. "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Journal of Evolution and Technology* 9, no. 1 (2002).

Brooke, John Hedley. "Visions of Perfectibility," *Journal of Evolution and Technology* 14, no. 2 (August 2005): 1-12.

Breazeal, Cynthia. *Designing Sociable Robots*. Cambridge, MA: MIT Press, 2002.

Brighton, Henry and Howard Selina. *Introducing Artificial Intelligence.* Thriplow, UK: Totem Books, 2003.

Brock, Brian. *Christian Ethics in a Technological Age*. Grand Rapids: Eerdmans Publishing, 2010.

Brooks, Rodney. *Flesh and Machines: How Robots Will Change Us*. New York: Vintage, 2003.

Brown, Warren S., Nancey Murphy, and H. Newton Malony, eds. *Whatever Happened to the Soul? Scientific and Theological Portraits of Human Nature.* Minneapolis: Fortress Press, 1998.

Caldera, Eva Orlebeke. "Cognitive Enhancement and Theories of Justice: Contemplating the Malleability of the Self," *Journal of Evolution and Technology* 18, no. 1 (May 2008): 116-123.

Calvin, John. *Institutes of the Christian Religion*. Edited by John T. McNeill. Louisville, KY: Westminster John Knox Press, 1960.

Cannon, Katie G. *Black Womanist Ethics*. Atlanta: American Academy of Religion, 1988.

Cannon, Katie Geneva. *Katie's Canon: Womanism and the Soul of the Black Community*. New York: Continuum Press, 1995.

Capek, Karel. *RUR (Rossum's Universal Robots)*. New York: Penguin Classics, 2004.

Card, Orson Scott. *Ender's Game*. New York: Starscape, 2002.

Carnagey, N. L. et al. "The effect of video game violence on physiological desensitization to real-life violence," *Journal of Experimental Social Psychology* 43, no. 3 (May 2007): 489-496.

Caruso, Denise. *Intervention: Confronting the Real Risks of Genetic Engineering and Life on a Biotech Planet.* San Francisco: Hybrid Vigor, 2006.

Cawsey, Alison. *The Essence of Artificial Intelligence.* New York: Prentice Hall, 1998.

Chapman, Audrey R. *Unprecedented Choices: Religious Ethics at the Frontiers of Genetic Science.* Minneapolis: Fortress Press, 1999.

Clayton, Philip. *Mind and Emergence*. New York: Oxford University Press, 2004.

Cole-Turner, Ronald, ed. *Design and Destiny: Jewish and Christian Perspectives on Human Germline Modification*. Cambridge, MA: MIT Press, 2008.

Colson, Charles W. and Nigel M. de S. Cameron, eds. *Human Dignity in the Biotech Century: A Christian Vision for Public Policy*. Downers Grove, IL: InterVarsity Press, 2004.

Conway Morris, Simon. *Life's Solution: Inevitable Humans in a Lonely Universe.* New York: Cambridge University Press, 2003.

Daly, Todd. "Life-Extension in Transhumanist and Christian perspectives: Consonance and Conflict," *Journal of Evolution and Technology* 14, no. 2 (August 2005): 57-75.

Darwin, Charles. *The Descent of Man.* New York: Penguin, 2004.

DeLashmutt, Michael W. "A Better Life Through Information Technology? The Techno-Theological Eschatology of Posthuman Speculative Science," *Zygon* 41, no. 2 (June 2006): 267-287.

Descartes, Rene. *Meditations, Objections, and Replies*. Edited and translated by Roger Ariew and Donald Cress. Indianapolis: Hackett Publishing, 2006.

De Waal, Frans and Peter L. Tyack, eds. *Animal Social Complexity: Intelligence, Culture, and Individualized Societies*. Cambridge, MA: Harvard University Press, 2003.

De Waal, Frans. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press, 1996.

De Waal, Frans. *Peacemaking Among Primates*. Cambridge, MA: Harvard University Press, 1989.

De Waal, Frans and Pier Francesco Ferrari. *The Primate Mind: Built to Connect With Other Minds*. Cambridge, MA: Harvard University Press, 2012.

Diana, Alison. "12 Advances in Medical Robotics," *Information Week Healthcare*, January 29, 2011. http://www.informationweek.com/news/galleries/healthcare/patient/229100383.

Dvorsky, George. "Better Living Through Transhumanism," *Journal of Evolution and Technology* 19, no. 1 (September 2008): 62-66.

Editors of Scientific American. *Understanding Artificial Intelligence*. New York: Warner Books, 2002.

Eiesland, Nancy. *The Disabled God: Toward a Liberatory Theology of Disability*. Nashville: Abingdon Press, 1994.

Evans, Woody. "Singularity Warfare: A Bibliometric Survey of Militarized Transhumanism," *Journal of Evolution and Technology* 16, no. 1 (June 2007): 161-165.

Fletcher, Joseph. *Humanhood: Essays in Biomedical Ethics*. Buffalo, NY: Prometheus, 1979.

Foerst, Anne. "Artificial Intelligence: Walking the Boundary," *Zygon* 31, no. 4 (December 1996): 681-693.

Foerst, Anne. *God in the Machine: What Robots Teach Us About Humanity and God*. New York: Dutton, 2004.

Fukuyama, Francis. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. New York: Farrar, Straus and Giroux, 2002.

Gardner, Howard. *Frames of Mind: The Theory of Multiple Intelligences*, 3rd edition. New York: Basic Books, 2011.

Garner, Stephen. "Transhumanism and Christian Social Concern," *Journal of Evolution and Technology* 14, no. 2 (August 2005): 29-43.

Geraci, Robert M. "Apocalyptic AI: Religion and the Promise of Artificial Intelligence," *Journal of the American Academy of Religion* 76, no. 1 (March 2008): 138-166.

Geraci, Robert. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. New York: Oxford University Press, 2010.

Gibson, William. *Neuromancer*. New York: Ace Books, 1984.

Gilson, Dave and Carolyn Perot. "It's the Inequality, Stupid: Eleven charts that explain what's wrong with America," *Mother Jones* (March/April 2011). http://motherjones.com/politics/2011/02/income-inequality-in-america-chart-graph.

Gould, Stephen Jay. *Ever Since Darwin: Reflections in Natural History*. New York: W. W. Norton, 1977.

Gould, Stephen Jay. *Full House: The Spread of Excellence from Plato to Darwin*. New York: Harmony Books, 1996.

Gould, Stephen Jay. *The Mismeasure of Man*. New York: W. W. Norton, 1996.

Gould, Stephen Jay. *Wonderful Life: The Burgess Shale and the Nature of History*. New York: W. W. Norton, 1989.

Gray, Chris Hables. *Cyborg Citizen: Politics in the Posthuman Age.* New York: Routledge, 2001.

Greenemeier, Larry. "Robots Arrive at Fukushima Nuclear Site With Unclear Mission," *Scientific American*, March 24, 2011. http://www.scientificamerican.com/article.cfm?id=robots-arrive-fukushima-nuclear.

Gregory of Nyssa. *On the Soul and the Resurrection*. Translated by Catherine P. Roth. Crestwood, NY: St. Vladimir's Seminary Press, 1993.

Hall, J. Storrs. *Beyond AI: Creating the Conscience of the Machine*. Amherst, NY: Prometheus, 2007.

Halley, Drew. "A Robot Stole My Job: Automation in the Recession," *Singularity Hub*, December 15, 2010. http://singularityhub.com/2010/12/15/a-robot-stole-my-job-automation-in-the-recession/.

Hambling, David. "Next-Gen Coal Mining Rescue Robot," *Popular Mechanics*, August 23, 2010. http://www.popularmechanics.com/science/energy/coal-oil-gas/next-gen-coal-mining-rescue-robot.

Hanks, Craig. *Technology and Values: Essential Readings*. Malden, MA: Wiley-

Blackwell, 2010.

Hanson, Robin. "Is a singularity just around the corner? What it takes to get explosive economic growth," *Journal of Transhumanism* 2 (June 1998).

Hardy, Edward R., ed. *Christology of the Later Fathers*. Louisville: KY: Westminster John Knox Press, 1954.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press, 1999.

Hefner, Philip. "The Animal that Aspires to be an Angel: The Challenge of Transhumanism," *Dialog: A Journal of Theology* 48, no.2 (June 2009): 158-167.

Hefner, Philip. *The Human Factor: Evolution, Culture, and Religion*. Minneapolis: Augsburg Fortress, 1993.

Hefner, Philip. *Technology and Human Becoming*. Minneapolis: Fortress Press, 2003.

Heller, Michael. *The Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation, and Costs Lives.* New York: Basic Books, 2008.

Herzfeld, Noreen. "Creating in Our Own Image: Artificial Intelligence and the Image of God," *Zygon* 37, no. 2 (June 2002): 303-316.

Herzfeld, Noreen. *In Our Image: Artificial Intelligence and the Human Spirit*. Minneapolis: Fortress Press, 2002.

Herzfeld, Noreen. *Technology and Religion: Remaining Human in a Co-Created World*. Conshohocken, PA: Templeton Press, 2009.

Hochschild, Jennifer L. *What's Fair? American Beliefs about Distributive Justice*. Cambridge, MA: Harvard University Press, 1981.

Honda Corporation. "History of ASIMO." http://asimo.honda.com/asimo-history/.

Hopkins, Patrick D. "A Moral Vision for Transhumanism," *Journal of Evolution and Technology* 19, no. 1 (September 2008): 1-5.

Hopkins, Patrick D. "Transcending the Animal: How Transhumanism and Religion Are and Are Not Alike," *Journal of Evolution and Technology* 14, no. 2 (August 2005): 13-28.

Hughes, James. *Citizen Cyborg: How Democratic Societies Must Respond to the*

Redesigned Humans of the Future. New York: Basic Books, 2004.

IEEE Intelligent Systems 21, no. 4 (July/August 2006).

iRobot Corporation. "iRobot Corporation: Our History." http://www.irobot.com/sp.cfm?pageid=203.

Jackelen, Antje. "The Image of God as Techno Sapiens," *Zygon* 37, no. 2 (June 2002): 289-302.

Jackson, Joseph. "The Amorality of Preference: A Response to the Enemies of Enhancement," *Journal of Evolution and Technology* 19, no. 1 (September 2008): 1-9.

Jones, Serene and Paul Lakeland, eds. *Constructive Theology: A Contemporary Approach to Classical Themes, with CD-ROM.* Minneapolis: Fortress Press, 2005.

Jones, Serene. *Feminist Theory and Christian Theology: Cartographies of Grace*. Minneapolis: Fortress Press, 1999.

Jordan, Gregory E. "Apologia for Transhumanist Religion," *Journal of Evolution and Technology* 15, no. 1 (February 2006): 55-72.

Kalil, Tom and Chuck Thorpe. "Supporting the President's National Robotics Initiative," White House Office of Science and Technology Policy, August 3, 2011. http://www.whitehouse.gov/blog/2011/08/03/supporting-president-s-national-robotics-initiative.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Edited by Mary Gregor. New York: Cambridge University Press, 1997.

Kaufman, Stuart. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. New York: Oxford University Press, 1995.

Keller, Catherine. *Apocalypse Now and Then: A Feminist Guide to the End of the World*. Minneapolis: Fortress Press, 2004.

Keller, Catherine. *Face of the Deep: A Theology of Becoming*. New York: Routledge, 2003.

Kenyon, Samuel H. "Would You Still Love Me If I Was A Robot?" *Journal of Evolution and Technology* 19, no. 1 (September 2008): 17-27.

Kierkegaard, Soren. *The Concept of Anxiety: A Simple Psychologically Orienting*

Deliberation on the Dogmatic Issue of Hereditary Sin. Edited and translated by Reidar Thomte and Albery B. Anderson. Princeton: Princeton University Press, 1980.

Kilner, John F., ed. *Why the Church Needs Bioethics: A Guide to Wise Engagement with Life's Challenges*. Grand Rapids: Zondervan, 2011.

Kilner, John F., Nigel M. de S. Cameron, and David L. Schiedermayer, eds. *Bioethics and the Future of Medicine: A Christian Appraisal*. Grand Rapids: Eerdmans Publishing, 1995.

Kilner, John F., Paige C. Cunningham, and W. David Hager, eds. *The Reproduction Revolution: A Christian Appraisal of Sexuality, Reproductive Technologies, and the Family*. Grand Rapids: Eerdmans Press, 2000.

Kilner, John F., Rebecca D. Pentz, and Frank E. Young, eds. *Genetic Ethics: Do the Ends Justify the Genes?* Grand Rapids: Eerdmans Publishing, 1997.

Kirby, Vicki. *Telling Flesh: The Substance of the Corporeal*. New York: Routledge, 1997.

Krueger, Oliver. "Gnosis in Cyberspace? Body, Mind and Progress in Posthumanism," *Journal of Evolution and Technology* 14 (August 2005): 77-89.

Kurzweil, Raymond. *The 10% Solution for a Healthy Life: How to Reduce Fat in Your Diet and Eliminate Virtually All Risk of Heart Disease and Cancer*. New York: Three Rivers Press, 1994.

Kurzweil, Ray. *The Age of Intelligent Machines*. Cambridge, MA: MIT Press, 1990.

Kurzweil, Ray. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking, 1999.

Kurzweil, Ray. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking, 2005.

Kurzweil, Ray and Terry Grossman. *Fantastic Voyage: Live Long Enough to Live Forever*. New York: Plume, 2005.

Kurzweil, Ray and Terry Grossman. *Transcend: Nine Steps to Living Well Forever*. New York: Rodale, 2010.

LaHaye, Tim and Jerry B. Jenkins. *Left Behind: A Novel of Earth's Last Days*. Carol Stream, IL: Tyndale House Publishers, 1995.

Lakoff, George and Mark Johnson. *Philosophy in the Flesh: The Embodied Mind and Its*

Challenge to Western Thought. New York: Basic Books, 1999.

LaTorra, Michael. "Trans-Spirit: Religion, Spirituality and Transhumanism," *Journal of Evolution and Technology* 15 (August 2005): 41-55.

LeVay, Simon. *When Science Goes Wrong: Twelve Tales from the Dark Side of Discovery*. New York: Plume, 2008.

Levy, David. *Love and Sex With Robots: The Evolution of Human-Robot Relationships*. New York: Harper Perennial, 2008.

Lossky, Vladimir. *In the Image and Likeness of God*. Crestwood, NY: St. Vladimir's Seminary Press, 1974.

MacIntyre, Alisdair. *After Virtue, 2nd Edition.* Notre Dame, IN: University of Notre Dame, 1984.

MacIntyre, Alisdair. *Dependent Rational Animals: Why Human Beings Need the Virtues*. Peru, IL: Open Court Press, 1999.

Maher, Derek F. and Calvin Mercer, eds., *Religion and the Implications of Radical Life Extension*. New York: Palgrave Macmillan, 2009.

Maienschein, Jane. *Whose View of Life? Embryos, Cloning, and Stem Cells*. Cambridge, MA: Harvard University Press, 2003.

Manoj, V. R. "Spiritual Transcendence in Transhumanism," *Journal of Evolution and Technology* 17, no. 1 (March 2008): 36-44.

Margulis, Lynn. *Symbiotic Planet: A New Look at Evolution*. New York: Basic Books, 1998.

McFadyen, Alistair. *Bound to Sin: Abuse, Holocaust and the Christian Doctrine of Sin.* New York: Cambridge University Press, 2000.

McFarland, Ian A. *Difference and Identity: A Theological Anthropology.* Cleveland: Pilgrim Press, 2001.

McFarland, Ian A. *The Divine Image: Envisioning the Invisible God*. Minneapolis: Fortress Press, 2005.

McKibben, Bill. *Enough: Staying Human in an Engineered Age*. New York: Times Books, 2003.

Menzel, Peter and Faith D'Aluisio. *Robosapiens: Evolution of a New Species.* Cambridge, MA: MIT Press, 2001.

Midgley, Mary. *Science as Salvation: A Modern Myth and Its Meaning*. New York: Routledge, 1992.

Minsky, Marvin. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon and Schuster, 2006.

Minsky, Marvin. *The Society of Mind*. New York: Simon and Schuster, 1986.

Mitchell, C. Ben, Robert D. Orr, and Susan A. Salladay, eds. *Aging, Death, and the Quest for Immortality*. Grand Rapids: Eerdmans Press, 2004.

Mitchell, C. Ben, et al., eds. *Biotechnology and the Human Good*. Washington, DC: Georgetown University Press, 2007.

Mithen, Steven. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Cambridge, MA: Harvard University Press, 2006.

Moltmann, Jurgen. *Science and Wisdom*. Translated by Margaret Kohl. Minneapolis: Fortress Press, 2003.

Moravec, Hans. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, MA: Harvard University Press, 1988.

Moravec, Hans. *Robot: Mere Machine to Transcendent Mind*. New York: Oxford University Press, 1999.

Moravec, Hans. "When will computer hardware match the human brain?" *Journal of Transhumanism* 1 (1998).

Murphy, Nancey. *Bodies and Souls, or Spirited Bodies?* New York: Cambridge University Press, 2006.

NASA Jet Propulsion Laboratory. http://www-robotics.jpl.nasa.gov/applications/index.cfm.

National Robotics Initiative, http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503641&org=CISE and http://www.nsf.gov/pubs/2011/nsf11553/nsf11553.htm.

Niebuhr, Reinhold. *The Nature and Destiny of Man*. Louisville, KY: Westminster John Knox Press, 1996.

Noble, David. *The Religion of Technology: The Divinity of Man and the Spirit of Invention*. New York: Alfred A. Knopf, 1997.

O'Donovan, Oliver. *Begotten or Made?* New York: Oxford University Press, 1984.

O'Hear, Anthony. *Beyond Evolution: Human Nature and the Limits of Evolutionary Explanation.* New York: Oxford University Press, 2002.

Okin, Susan Moller. *Justice, Gender, and the Family*. New York: Basic Books, 1989.

Peacocke, Arthur. *Theology for a Scientific Age: Being and Becoming - Natural, Divine, and Human*. Minneapolis: Fortress Press, 1993.

Pegis, Anton C., ed. *Basic Writings of Thomas Aquinas, Vol. 1*. Indianapolis: Hackett Publishing, 1997.

Perkowitz, Sidney. *Digital People: From Bionic Humans to Androids.* Washington, DC: Joseph Henry Press, 2004.

Peterson, James C. *Changing Human Nature: Ecology, Ethics, Genes, and God.* Grand Rapids: Eerdmans Publishing, 2010.

Placher, William C., ed. *Essentials of Christian Theology*. Louisville, KY: Westminster John Knox Press, 2003.

President's Council on Bioethics. *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. Washington, DC: US Government Printing Office, 2003.

Ramsey, Paul. *Ethics at the Edges of Life: Medical and Legal Intersections*. New Haven, CT: Yale University Press, 1978.

Ramsey, Paul. *The Patient as Person: Explorations in Medical Ethics*. New Haven, CT: Yale University Press, 1970.

Rawls, John. *A Theory of Justice, Revised Edition*. Cambridge, MA: Belknap Press, 1999.

Rolnick, Philip A. *Person, Grace, and God.* Grand Rapids: Eerdmans Publishing, 2007.

Rose, Nikolas. *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*. Princeton: Princeton University Press, 2007.

Rossano, Matt J. "Artificial Intelligence, Religion, and Community Concern," *Zygon* 36, no. 1 (March 2001): 57-75.

Rothman, Stephen. *Life Beyond Molecules and Genes: How Our Adaptations Make Us Alive*. Conshohocken, PA: Templeton Press, 2009.

Saenz, Aaron. "Robotic Labor Taking Over the World? You Bet - Here Are the Details," *Singularity Hub*, September 12, 2011. http://singularityhub.com/2011/09/12/robotic-labor-taking-over-the-world-you-bet-here-are-the-details/.

Schleiermacher, Friedrich. *The Christian Faith*. Edited by H. R. Mackintosh and J. S. Stewart. New York: T&T Clark, 1999.

Sex Robot, Discovery Fit and Health, 2011.

Shannon, C. E. "A mathematical theory of communication," *Bell System Technical Journal* 27 (July and October 1948): 379-423, 623-656.

Shults, F. LeRon and Brent Waters, eds. *Christology and Ethics*. Grand Rapids: Eerdmans Publishing, 2010.

Shults, F. LeRon, ed. *The Evolution of Rationality: Interdisciplinary Essays in Honor of J. Wentzel vanHuyssteen.* Grand Rapids: Eerdmans Press, 2006.

Shults, F. LeRon. *Reforming Theological Anthropology: After the Philosophical Turn to Relationality*. Grand Rapids: Eerdmans Press, 2003.

Singer, P.W. *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. New York: Penguin, 2009.

Singularity University. "Partners." http://singularityu.org/about/partners/.

Sorgner, Stefan Lorenz. "Beyond Humanism: Reflections on Trans- and Posthumanism," *Journal of Evolution and Technology* 21, no. 2 (October 2010): 1-19.

Tamatea, Laurence. "If robots R-US, who am I: Online 'Christian' responses to artificial intelligence," *Culture and Religion* 9, no. 2 (July 2008): 141-160.

Tanner, Kathryn. *Christ the Key*. New York: Cambridge University Press, 2010.

Tanner, Kathryn. *Jesus, Humanity, and the Trinity: A Brief Systematic Theology*. Minneapolis: Fortress Press, 2001.

Tanner, Kathryn. *The Politics of God: Christian Theologies and Social Justice*. Minneapolis: Fortress Press, 1992.

Taylor, Carol R. and Roberto Dell'Oro, eds. *Health and Human Flourishing: Religion, Medicine, and Moral Anthropology*. Washington, DC: Georgetown University Press, 2006.

Toyota Motor Corporation. "TMC Shows New Nursing and Healthcare Robots in Tokyo." November 1, 2011. http://pressroom.toyota.com/releases/tmc+shows+new+nursing+healthcare+robots+tokyo.htm.

Troeltsch, Ernst. *The Social Teaching of the Christian Churches*. Louisville, KY: Westminster John Knox Press, 1992.

Turkle, Sherry, ed. *The Inner History of Devices.* Cambridge, MA: MIT Press, 2008.

Turkle, Sherry. *Life on the Screen: Identity in the Age of the Internet.* New York: Simon and Schuster, 1995.

Turkle, Sherry. *The Second Self: Computers and the Human Spirit, Twentieth Anniversary Edition*. Cambridge, MA: MIT Press, 2005.

United Nations. *World Population Prospects, 2004 Revision, Volume III: The Analytical Report*. United Nations, 2005.

US Defense Advanced Research Projects Agency. http://www.darpa.mil/NewsEvents/Budget.aspx.

US Department of Defense. "Program Announcement: Department of Defense (DoD) Fiscal Year 2012 Defense University Research Instrumentation Program (DURIP). http://www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2011/BAA-AFOSR-DURIP-2011.ashx.

US National Institute of Standards and Technology. http://www.nist.gov/el/isd/robotics.cfm and http://www.nist.gov/el/isd/ks/response_robot_test_methods.cfm

US Robotics. "A Roadmap for US Robotics - From Internet to Robotics," May 21, 2009, http://www.us-robotics.us/reports/CCC%20Report.pdf.

VanHuyssteen, J. Wentzel. *Alone in the World? Human Uniqueness in Science and*

Theology. Grand Rapids: Eerdmans Press, 2006.

Verdoux, Philippe. "Transhumanism, Progress, and the Future," *Journal of Evolution and Technology* 20, no. 2 (2009): 49-69.

Vinge, Vernor. "The Coming Technological Singularity: How to Survive in the Post-Human Era," VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, March 30-31, 1993.

Vinge, Vernor. *Marooned in Realtime.* New York: Baen, 1986.

Wallach, Wendell and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press, 2009.

Waters, Brent. *From Human to Posthuman: Christian Theology and Technology in a Postmodern World*. Burlington, VT: Ashgate Publishing, 2006.

Waters, Brent. *Reproductive Technology: Towards a Theology of Procreative Stewardship*. Cleveland: Pilgrim Press, 2001.

Waters, Brent. *This Mortal Flesh: Incarnation and Bioethics*. Grand Rapids: Brazos Press, 2009.

Waters, Brent and Ronald Cole-Turner, eds. *God and the Embryo: Religious Voices on Stem Cells and Cloning.* Washington, DC: Georgetown University Press, 2003.

West, Traci C. *Disruptive Christian Ethics: When Racism and Women's Lives Matter*. Louisville, KY: Westminster John Knox Press, 2006.

Woolston, Chris. "Robots built to help autistic children," *Los Angeles Times*, October 17, 2011. http://articles.latimes.com/2011/oct/17/health/la-he-autism-robots-20111017.

World Health Organization. "WHO World Malaria Report 2010," World Health Organization. http://www.who.int/malaria/world_malaria_report_2010/malaria2010_summary_keypoints_en.pdf.

Wright, Robert. *Nonzero: The Logic of Human Destiny*. New York: Vintage Books, 2000.

Young, Jeffrey. "Programmed for Love: In a skeptical turn, the MIT ethnographer Sherry Turkle warns of the dangers of social technology," *Chronicle of Higher Education*, January 14, 2011. http://chronicle.com/article/Programmed-for-Love-The/125922.

Zizioulas, John D. *Being as Communion: Studies in Personhood and the Church*. Crestwood, NY: St. Vladimir's Seminary Press, 1997.